



```

Db      758 GACGCTATAGAGTTTCATTATAAATTTTACAAAATAAAATCAGCAAAATAATTTTTC 817

QY      241 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 274
      |||||
Db      818 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 851

RESULT 2
AB023041/c      83650 bp      DNA      linear      PLN 14-FEB-2004
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MPE11.
DEFINITION      AB023041 BA000014
ACCESSION      AB023041.1 GI:4220640
VERSION
KEYWORDS
SOURCE
  ORGANISM
    Arabidopsis thaliana (thale cress)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
  AUTHORS      Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
  TITLE      Structural analysis of Arabidopsis thaliana chromosome 3. I.
  TITLE      Sequence features of the regions of 4,504,864 bp covered by sixty
  P1 and TAC clones
  JOURNAL      DNA Res. 7 (2), 131-135 (2000)
  PUBMED      10819329
  REFERENCE    2 (bases 1 to 83650)
  AUTHORS      Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
  TITLE      Direct Submission
  JOURNAL      Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research
  Institute, Department of Plant Gene Research, 1532-3, Yana,
  Kisarazu, Chiba 252-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
  Tel:81-438-52-3935, Fax:81-438-52-3934)
  COMMENT      Address for correspondence: kaos@kazusa.or.jp
  For the latest information on annotation of this clone, please see
  http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MPE11
  Genes with similarity to proteins in the databases are described in
  'product' or 'note' qualifiers. Genes that have no significant
  protein similarity are described as 'unknown protein'.
  The software programs used to predict genes include: Grail
  (Informatics Group, Oak Ridge National Laboratory,
  http://combio.ornl.gov/Grail-1.3/),
  GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
  NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
  Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
  SplicePredictor (Volker Brendel, Stanford University,
  http://gremlin.zool.iastate.edu/cgi-bin/sp.cgi).
  Genes encoding tRNAs are predicted by tRNAscan-SE
  (Sean Eddy, Washington University School of Medicine, St. Louis,
  http://genome.wustl.edu/eddy/tRNAscan-SE/).
  This sequence may not be the entire insert of this clone. It may be
  shorter because we remove overlaps between neighboring submissions.
  The 5' clone is K9122 and the 3' clone is MJL14.

FEATURES             Location/Qualifiers
     source           1..83650
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosomes="3"
                     /clone="MPE11"
                     /clone_lib="Miteui P1"
                     /ecotype="Columbia"
                     /complement(542..764)
                     /note="CDS is reported in Acc# AP000599
                     contains similarity to CYP-rich zinc finger protein
                     gene id:K9122.5"
                     /number=1
                     /evidence=not experimental
                     join(1705..2463,2548..2739,2862..2949,3037..3107,
                     3190..3273,3410..3514)
                     /note="gb|AAD55139.1
                     gene id:MPE11.1"
                     /codon_start=1

exon
CDS
  complement(11450..11818)
  /note="unnamed protein product; gene_id:MPE11.4
  unknown protein"
  /codon_start=1
  /evidence=not experimental
  /protein_id="BAB01050.1"
  /db_xref="GI:9279592"
  /translation="MNSLKKEERVEEDNGKSDGNRGKPKSTEVVTVTEEEVDEPFKI
  LRRHVATRTVAKVNGVGAEGELPSKKRSQNLGLRSLDCNGVDRGDFEINRVGL
  GGLGDLNCKPEPDSVSUSL"
  16981..17979
  /note="gene id:MPE11.5"
  /codon_start=1
  /evidence=not experimental
  /product="AP2 domain transcription factor-like protein"
  /protein_id="BAB01051.1"
  /db_xref="GI:9279593"
  /translation="MAERKGRSSIQTNKPNKKPKPKPQLMHLPLSBDLKTMRKLR
  FVNDPVATDYSSSEERSSRRKRYVCEIDLPPFAAATQAESSESYCOESNNNGVSK
  TKISACKVLRSKASPVVGRSVTVKPVGRQKWKGAARIRHPIKTVTWLQY
  ETLQADADAYATKLEFLDALAATAASAASVLSNESGSMLSAGSSIDLDKKLVDSTLD
  QQAGSKASDFDFADLQIPMGCFIDDSFIPNACGLDFLITEENNMLDDYCGID
  DLDITGLECDGSELPDYDFSDVEIDLGITTDKYAPVDVHIATTTPTPLNACP"
  join(21893..22063,22330..22386,22578..22783,22894..23227,
  23311..23382,23473..23613)
  /note="gb|RAF23821.1
  gene_id:MPE11.6"
  /codon_start=1
  /evidence=not experimental
  /product="homocysteine S-methyltransferase AtHMT-1"
  /protein_id="BAB01052.1"
  /db_xref="GI:9279594"
  /translation="MVLEKSALEDLLKCGCAVVDGGFATQLEIGHAINDPLMS
  AVSLKNPELIRKVMETLEAGADIVTVSSYQATIPGFSRGLSTEESSLQKSVEL
  AYAEARDFWEKSVKSYNSYNALVAASIGSGAYLADSEYSGHGVNSGLDKLQDF
  HRRRLQVLVEAGFDLLAFETIPNLEAQACVBLEEEKVQIPAWCTFSDVDEKAPSG
  ESFEECLEPLNKNNIYAVGNCAPPQFIENIRKFAKLTKEAIVVYPNSGEYWDGKA
  KQWLPSQCFDDEFEFMATKWRDLGAKLIGGCCRTTPTSTINAISRDLKER"

```

	Query Match	99.4%	Score 272.4	DB 15	Length 83650	
	Best Local Similarity	99.6%	Pred. No. 5.8e-52			
	Matches 273	Conservative	0	Mismatches	0	Gaps 0
Qy	1	TCTGAAGAATAGCCGATACCAATTTTTCCACATCGAGCTTCTTTATTCGAAAGTCA	60			
Db	12103	TCTAAAATAGCCGATACCAATTTTTCCACATCGAGCTTCTTTATTCGAAAGTCA	12044			
Qy	61	ATAAAGTGTGACGTCATGACTCTTACGCTTTAAACATCGCATGATGATGCTCAATTAGCAT	120			



## AC010633/c

LOCUS AC010633 195931 bp DNA linear PRI 31-JUL-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2317K6, complete sequence.

ACCESSION AC010633  
 VERSION AC010633.8 GI:15042798  
 KEYWORDS HTG.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 TITLE Homnidae; Homo.

REFERENCE 1 (bases 1 to 195931)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Unpublished

REFERENCE 2 (bases 1 to 195931)  
 AUTHORS DOE Joint Genome Institute.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 195931)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 195931)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

COMMENT Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Jul 31, 2001 this sequence version replaced gi:13677000.  
 Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.6% of Sequence;

Estimated Total Number of Errors is 0.4.  
 Location/Qualifiers

1..195931  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2317K6"

ORIGIN  
 Query Match 16.6%; Score 45.4; DB 8; Length 195931;  
 Best Local Similarity 56.3%; Pred. No. 1;  
 Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 103 TGATGATGTCATTAGCATCAATCTCCACCGCTATAAGAGTTTCATTATAAATTTTACAAAATAAAA 222  
 Db 105243 TAATGAAGTAGGAGGCAAGTTATCCATTGAAGTTATTTATTTTGCAGCTCATCTTAA 105184

QY 163 CCGTCTAAGTTCCACACCGGCTATAAGAGTTTCATTATAAATTTTACAAAATAAAA 222  
 Db 105183 GTGACAAAATTCATACAGAGACTATACAGAAATCAATATTTAATATATATAAATAAATTAAT 105124

QY 223 TCAGCAAAATTAATTTTCTTGACTAAGCTTA 253  
 Db 105123 ACTTCAAAATATCTTTCACATTAAGATGATTA 105093

RESULT 9  
 AC073652 167034 bp DNA linear HTG 26-JUN-2001  
 LOCUS Homo sapiens chromosome 12 clone RP11-148D15, WORKING DRAFT

DEFINITION SEQUENCE.

ACCESSION AC073652  
 VERSION AC073652.15 GI:14547458

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 TITLE Homnidae; Homo.

REFERENCE 1 (bases 1 to 167034)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

## RESULT 7

AC008964/c

LOCUS AC008964 143841 bp DNA linear PRI 08-JUN-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2360A17, complete sequence.

ACCESSION AC008964  
 VERSION AC008964.6 GI:14329064

KEYWORDS HTG.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 TITLE Homnidae; Homo.

REFERENCE 1 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Unpublished

REFERENCE 2 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

REFERENCE 3 (bases 1 to 143841)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
 TITLE Direct Submission

COMMENT Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Jun 8, 2001 this sequence version replaced gi:13876485.  
 Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.9% of Sequence;

Estimated Total Number of Errors is 0.1.  
 STS Content:

SHGC-8813 GI4344.  
 Location/Qualifiers

1..143841  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2360A17"

ORIGIN  
 Query Match 16.6%; Score 45.4; DB 8; Length 143841;  
 Best Local Similarity 56.3%; Pred. No. 1.1;  
 Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 103 TGATGATGTCATTAGCATCAATCTCCACCGCTATAAGAGTTTCATTATAAATTTTACAAAATAAAA 222  
 Db 18293 TAATGAAGTAGGAGGCAAGTTATCCATTGAAGTTATTTATTTTGCAGCTCATCTTAA 18234

QY 163 CCGTCTAAGTTCCACACCGGCTATAAGAGTTTCATTATAAATTTTACAAAATAAAA 222  
 Db 18233 GTGACAAAATTCATACAGAGACTATACAGAAATCAATATTTAATATATATAAATAAATTAAT 18174

QY 223 TCAGCAAAATTAATTTTCTTGACTAAGCTTA 253  
 Db 18173 ACTTCAAAATATCTTTCACATTAAGATGATTA 18143

RESULT 8

Alebrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Lucier, A., Lucier, R., Luna, R., Luisleged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Snaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, D., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: HBPB  
Center clone name: RP11-148D15  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 66% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 168964 bases at least Q40  
Consensus quality: 170984 bases at least Q30  
Estimated insert size: 167033; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 167034: contig of 167034 bp in length.  
FEATURES  
    Location/Qualifiers  
        1..167034  
        /organism="Homo sapiens"  
        /mol\_type="genomic DNA"  
        /db\_xref="taxon:9606"  
        /chromosome="12"  
        /clone="RP11-148D15"  
ORIGIN  
    Query Match 16.1%; Score 44.2; DB 14; Length 167034;  
    Best Local Similarity 51.2%; Pred. No. 2;  
    Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
    48 ATTCAAAAGTCAATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATG 107  
    92792 ATTATCAAAATATACATAGAAATTTGATTTCTTAATAGCAATATACATCACATTTTG 92851  
    108 ATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTAGTTGTTGCAATATCGACCGTC 167  
    92852 AATTGATTTTTTAACCACTACACTATTTCAGTGTATATATGTGCAGAAACATTTATAAAG 92911  
    168 TTAGTTCCACCGACGGCTATAGAGTTTCATTAATAATTTTACCAAAATATAAATCAGC 227  
    92912 GATAGTTTCATCCATGAAGATGAAAAATTTGATCATGAAGATTTACTACTATAAAAGT 92971  
    228 AATAAATTTTTTCTTGACTAA 248  
    92972 CATTTCTTTTATTGTTTAA 92992  
LOCUS AC073528 167041 bp DNA linear PRI 01-MAY-2001  
DEFINITION Homo sapiens 12 BAC RP11-8P16 (Roswell Park Cancer Institute Human BAC Library) complete sequence.  
ACCESSION AC073528  
VERSION AC073528.26 GI:13899199  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 167041)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jolivet, S., Joudah, S., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Luisleged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador M., Mel, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K. T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunoye, G., Orsgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pull, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shm, C., Shoocharti, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Kucherlapati, R. and Gibbs, R.

## Direct Submission

Unpublished

[illegible]

Таблица 2

Worley, K. C.

Submitted (22-JUN-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 167041)

2019

Worley, K.C.

Submitted (01-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 1, 2001 this sequence version replaced qi:13876431.

On May 1, 2001, the following information was received from the following sources:

INFORMATION: ILCP:77

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

STSSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

**QUALITY OF INDIVIDUAL BASES:**This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at [URL:](#)

<http://qc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

CIVIL-STAT-REPORT-----

----- Summary Statistics -----

Contig length:	167603
Phrap values in estimate:	166580
Average error rate (BCM:Phrap estimate):	3.901e-06
Fraction of Phrap values less than 40 :	0.000210109
Number of consensus changing edits:	0
Number of N's in consensus :	0

Position	Consensus	changing edits	Edited+Context
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	1
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	1
19	1	1	1
20	1	1	1
21	1	1	1
22	1	1	1
23	1	1	1
24	1	1	1
25	1	1	1
26	1	1	1
27	1	1	1
28	1	1	1
29	1	1	1
30	1	1	1
31	1	1	1
32	1	1	1
33	1	1	1
34	1	1	1
35	1	1	1
36	1	1	1
37	1	1	1
38	1	1	1
39	1	1	1
40	1	1	1
41	1	1	1
42	1	1	1
43	1	1	1
44	1	1	1
45	1	1	1
46	1	1	1
47	1	1	1
48	1	1	1
49	1	1	1
50	1	1	1
51	1	1	1
52	1	1	1
53	1	1	1
54	1	1	1
55	1	1	1
56	1	1	1
57	1	1	1
58	1	1	1
59	1	1	1
60	1	1	1
61	1	1	1
62	1	1	1
63	1	1	1
64	1	1	1
65	1	1	1
66	1	1	1
67	1	1	1
68	1	1	1
69	1	1	1
70	1	1	1
71	1	1	1
72	1	1	1
73	1	1	1
74	1	1	1
75	1	1	1
76	1	1	1
77	1	1	1
78	1	1	1
79	1	1	1
80	1	1	1
81	1	1	1
82	1	1	1
83	1	1	1
84	1	1	1
85	1	1	1
86	1	1	1
87	1	1	1
88	1	1	1
89	1	1	1
90	1	1	1
91	1	1	1
92	1	1	1
93	1	1	1
94	1	1	1
95	1	1	1
96	1	1	1
97	1	1	1
98	1	1	1
99	1	1	1
100	1	1	1

----- Distribution of Quality &lt; 40 Bases

[illegible]

## FEATURES

```

Version: 1.01  qxf0.
Location/Qualifiers
1. .167041
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon.9606"
/chromosome="12"
/clone="RP11-8P16"
complement(1. .2013)
/note="Overlaps bases 2013. .1 of clone AC068800"
/function="Overlaps with adjacent clone AC068800"
complement(440. .488)
/rpt_family="L2"
complement(1346. .1552)
/rpt_family="MIR"
1765. .1795
/rpt_family="AT_rich"
2143. .2237
/rpt_family="MER112"
2624. .2650
/rpt_family="(TC)n"
2678. .2820
/rpt_family="L2"
3733. .3775
/rpt_family="(TAA)n"
3834. .3947
/rpt_family="L2"
complement(4126. .4246)
/rpt_family="MIR"
4335. .4383
/rpt_family="(TGAA)n"
5743. .5769
/rpt_family="(TTTTG)n"
complement(5770. .6059)
/rpt_family="AluSx"
6253. .6473
/rpt_family="MIR"
6687. .6719
/rpt_family="(TG)n"
7597. .7611
/rpt_family="AT_rich"
complement(7612. .7923)

```

```

/rpt_family="AluYa5"
7924..7941
/rpt_family="AT_rich"
complement(8256..8962)
/rpt_family="L2"
8966..9036
/rpt_family="MIR"
complement(9803..10024)
/rpt_family="L2"
11288..11321
/rpt_family="AT_rich"
11478..12092
/rpt_family="L2"
complement(12162..12303)
/rpt_family="MLT1G"
complement(12398..12636)
/rpt_family="MLT1G"
complement(12641..13107)

Query Match      16.1%; Score 44.2; DB 8; Length 167041;
Best Local Similarity 51.2%; Pred. No. 2;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 48 ATTCCAAAGTCAATAAGTGTGACGTGATGATCTAGCGTTTAAACATCGCATGATG 107
Db 74249 ATTATCAAAATATACACATAGAAATTTGATTTAAATAGCAATTAACATCACATTTTG 74190

QY 108 ATGTCAATAGCATCAATCTCCACCGTCCCAATTTTATTAGTTGTCACATATCGACCGTC 167
Db 74189 AATTGATTTTAAACCACTACACTATTCTAGTGTATATGTGCGAGGAACATTTATAAG 74130

QY 168 TAAAGTTCCACACCGGCTTAAGAGTTTCATTTAAATTTTAGCAAAATAAATCAGC 227
Db 74129 GATAGTTCAATCCCATGAAGATGAAATTTGATCATGAAGATTACTACTATAAAAGT 74070

QY 228 AAATAATTTTCTTGACTAA 248
Db 74069 CATTCTCTTTTATTGTTTAA 74049

RESULT 11
AC024628 207976 bp DNA linear HTG 25-MAY-2000
LOCUS Homo sapiens chromosome 12 clone RP11-632011 map 12, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
ACCESSION AC024628
VERSION AC024628.3 GI:8076668
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 207976)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Baetien,V., Beda,F.,
Boguslavsky,L., Bouckgaeter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneau,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,

```

```

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7144995.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5991
Center clone name: 632_O_11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 193853 bases at least Q40
Consensus quality: 200610 bases at least Q30
Consensus quality: 203401 bases at least Q20
Insert size: 228000; agarose-fp
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1209: contig of 1209 bp in length
* 1210 1309: gap of 100 bp
* 1310 2946: contig of 1637 bp in length
* 2947 3046: gap of 100 bp
* 3047 5028: contig of 1982 bp in length
* 5029 5128: gap of 100 bp
* 5129 7145: contig of 2017 bp in length
* 7146 7245: gap of 100 bp
* 7246 8721: contig of 1476 bp in length
* 8722 8821: gap of 100 bp
* 8822 11031: contig of 2210 bp in length
* 11032 11131: gap of 100 bp
* 11132 12855: contig of 1724 bp in length
* 12856 12955: gap of 100 bp
* 12956 15925: contig of 2970 bp in length
* 15926 16025: gap of 100 bp
* 16026 18959: contig of 2934 bp in length
* 18960 19059: gap of 100 bp
* 19060 22533: contig of 3474 bp in length
* 22534 27457: gap of 100 bp
* 27457 27566: gap of 100 bp
* 27566 29768: contig of 2212 bp in length
* 29769 29868: gap of 100 bp
* 29869 34812: contig of 4944 bp in length
* 34813 34913: gap of 100 bp
* 34913 41011: contig of 6039 bp in length
* 41012 41111: gap of 100 bp
* 41112 45236: contig of 4125 bp in length
* 45237 45337: gap of 100 bp
* 45337 55231: contig of 9894 bp in length
* 55231 55330: gap of 100 bp

```

TITLE  
JOURNAL  
COMMENT



gap 29769. 29868

Haugen, E., Gillett, W., Zhou, Y., James, R., Phelps, K., Iadanoto, S.,  
 Bub, K., Simms, E., Levy, R., Clendenning, J., Kaul, R., Kent, W.J.,  
 Furey, T.S., Baertsch, R.A., Brent, M.R., Keibler, E., Flicek, P.,  
 Bork, P., Suyama, M., Bailey, J.A., Portnoy, M.E., Torrents, D.,  
 Chinwalla, A.T., Gish, W.R., Eddy, S.R., McPherson, J.D., Olson, M.V.,  
 Eichler, E.E., Green, E.D., Waterston, R.H. and Wilson, R.K.  
 The DNA sequence of human chromosome 7  
 Nature 424 (6945), 157-164 (2003)  
 12853948  
 2 (bases 1 to 266344)  
 Bauer, C., McPherson, C. and Williams, D.  
 The sequence of Homo sapiens BAC clone GS1-250N6  
 Unpublished (2001)  
 3 (bases 1 to 266344)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (20-JUN-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 4 (bases 1 to 266344)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (17-JUN-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 266344)  
 Waterston, R.  
 Direct Submission  
 Submitted (30-SEP-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 6 (bases 1 to 266344)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (25-MAR-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 7 (bases 1 to 266344)  
 Waterston, R.  
 Direct Submission  
 Submitted (04-JUN-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 8 (bases 1 to 266344)  
 Wilson, R.  
 Direct Submission  
 Submitted (30-JAN-2004) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Mar 25, 2001 this sequence version replaced gi:5091650.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_GS250N06  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and  
 sequencing collaboration between the NHGRI Chromosome 7 Mapping  
 Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington  
 University Genome Sequencing Center. For additional information  
 about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
 mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

SOURCE INFORMATION:  
 This clone is from the first BAC library from Genome Systems, Inc.  
 (<http://www.genomesystems.com>).

Cell line: lymphoblastoid  
 Haplotypes: two  
 VECTOR: pBeloBAC  
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
 Actual start of this clone is at base position 1 of GS1-250N6  
 actual end is at base position 266344 of GS1-250N6.

FEATURES	Source
repeat_region	1..266344 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="7" /map="7" /clone="GS1-250N6" /clone_lib="GSBAC1" 762..902 /rpt_family="MIR"
repeat_region	3681..3865 /rpt_family="MER1_type"
repeat_region	4657..4713 /rpt_family="(TA)n"
repeat_region	5001..5125 /rpt_family="MIR"
repeat_region	6297..6466 /rpt_family="MER1_type"
repeat_region	10202..10520 /rpt_family="Alu"
repeat_region	11275..11427 /rpt_family="MIR"
repeat_region	11592..11824 /rpt_family="MIR"
repeat_region	12004..12103 /rpt_family="Alu"
repeat_region	12899..13675 /rpt_family="L1"
repeat_region	14550..14874 /rpt_family="Alu"
repeat_region	16275..16580 /rpt_family="Alu"
repeat_region	18336..18380 /rpt_family="CT-rich"
repeat_region	19426..19681 /rpt_family="MIR"
repeat_region	19691..19827 /rpt_family="L2"
repeat_region	19857..20139 /rpt_family="L2"
repeat_region	20146..20482 /rpt_family="MER2_type"
repeat_region	20483..20629 /rpt_family="L2"
repeat_region	21873..22161 /rpt_family="Alu"
repeat_region	22458..22842 /rpt_family="L2"
repeat_region	23395..23447 /rpt_family="GA-rich"
repeat_region	23875..24166 /rpt_family="Alu"
repeat_region	25208..25431 /rpt_family="Alu"
repeat_region	28192..28253

```

repeat_region /rpt_family="GA-rich"
28282..28508
repeat_region /rpt_family="MIR"
30611..30652
gene /rpt_family="(CA)n"
complement(30701..105771)
/gene="GLI3"
complement(join(30701..30943,105605..105771))
/gene="GLI3"
complement(join(<30701..30943,105605..105728))
/gene="GLI3"
/note="Homo sapiens GLI-Kruppel family member GLI3 (Greig
cephalopolysyndactyly syndrome) (GLI3), mRNA.;
H.GS250N06.1
This gene was based on gi(13518031)
Continues as H.GS489L14.1"
/codon_start=1
/product="unknown"
/protein_id="AA02015.1"
/db_xref="GI:41393490"
/translation="MRAQSHSTTTEKKKVENIVKCSRTDVSSEKAVASSTTSNEDS
SPGQTHRRNRNITMPQNVQGLSKVSEPTSSDERASLLIKKEIHGSLPHVAPSV
PYRGTVFAMDPRNGYMEPHY"
repeat_region 34549..34579
/rpt_family="(TTTTG)n"
repeat_region 34931..34953
/rpt_family="(TCTCCC)n"
repeat_region 36317..36368

Query Match 15.8%; Score 43.4; DB 8; Length 266344;
Best Local Similarity 50.2%; Pred. No. 2.7;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 ATTTTCCACATGCGATCTCTTTATTCACAAAGTCATAAAGTGTGACGTCATGATA 81
Db 162852 AAAATTATCTGAATATACTTTCCTTATTGAGAATTTCCAAAATTAAGATCTCAAAAAC 162911

QY 82 CTTACGCTTTAAACATCCGATCATGTCATTAGCATCAATCTCCACGTCCTCAATTTA 141
Db 162912 AGCATAGTTAGTAGAAGAATAAATATCATTTACAGATGAATTTCTAATGTCAAAAATA 162971

QY 142 TTTAGTTGTTGACATATCACCCTCTAAGTTCCACACCGCGCTATAGAGTTTTCATT 201
Db 162972 ATTTCTAAAGATTTCCACAAATGAAACACATAATATCCGCTTGAAACCTCAIT 163031

QY 202 ATAAATTTAGCAAAATAAATAACAGCAATAAT 234
Db 163032 TTATATTTGCCAAAATAAGAACAGAGGAAT 163064

RESULT 13
AC008563
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTC-536A23, linear HTG 18-JUL-2000
7 ordered pieces.
AC008563
VERSION AC008563.4 GI:7711299
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 94703)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
2 (bases 1 to 94703)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 6, 2000 this sequence version replaced gi.7708951.

```

```

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 391027, H433
Center clone name: CIT-HSPC_536A23
-----
Summary Statistics
Consensus quality: 88918 bases at least Q40
Consensus quality: 93000 bases at least Q30
Consensus quality: 93760 bases at least Q20
Estimated insert size: 99000; pulse field gel estimation
Estimated insert size: 94453; sum-of-contigs estimation
Quality coverage: 5.84 in Q20 bases; pulse field gel estimation
Quality coverage: 6.12 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 34594: contig of 34594 bp in length
* 34595 5143: gap of unknown length
* 34695 5143: contig of 16449 bp in length
* 51144 51243: gap of unknown length
* 51244 66425: contig of 15182 bp in length
* 66426 72366: gap of unknown length
* 66526 72366: contig of 5841 bp in length
* 72367 72467: gap of unknown length
* 72467 79229: contig of 7163 bp in length
* 79229 79729: gap of unknown length
* 79730 80679: contig of 950 bp in length
* 80680 80779: gap of unknown length
* 80780 94703: contig of 13924 bp in length.
FEATURES
source
1..94703
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-536A23"
/clone_lib="Caltech human BAC library C"
34595..34694
/estimated_length=unknown
51144..51243
/estimated_length=unknown
66426..66525
/estimated_length=unknown
72367..72466
/estimated_length=unknown
79630..79729
/estimated_length=unknown
80680..80779
/estimated_length=unknown
ORIGIN
Query Match 15.5%; Score 42.4; DB 14; Length 94703;
Best Local Similarity 48.4%; Pred. No. 5.8;
Matches 118; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 4 AGAATATATAGCCGATACCAATTTTCCACATGACCTTCCTTTATTCAAAAGTCAATA 63
Db 72000 AGTCTTAATACAAAGATTAATAATTTTAGATGGTGATATAATTTTCAAAAAGTAACC 72059
QY 64 AAGTGTGACGTCAATGATCTAGCTTTTAAACATCGCATGATGATGATGATGATGATGAT 123
Db 72060 AATGCTGGGGTTATTAGAAATTCGATTTAGTTACTGACAGGATGATATTATGATGCTG 72119

```

```
QY 124 TCTCCACCGTCCAAATTTATTTAGTTGTTGACAAATATCGACCGCTCTAAGTTCACACCGAC 183
Db 72120 TTTTCTAAGGTGAGTAAATTTCAATGAATATATCTTTTCCACCATTTTACAAATTTGAAAC 72179
QY 184 GCGTATAAGAGTTTCATTATATAATTTTACAAAAATAGCAATATAGCAATATATTTTCTTG 243
Db 72180 TGATATGAATTCATTTTACACCTTCTGAACTCACTGACGACGACGACGACGACGACGAC 72239
QY 244 ACTA 247
Db 72240 CATA 72243

RESULT 14
LOCUS AX209879/c 229 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1719 from Patent WO0157207.
ACCESSION AX209879
VERSION AX209879.1 GI:15424302
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1
AUTHORS Algate, P.A. and Mannion, J.
TITLE Compositions and methods for the therapy and diagnosis of ovarian
JOURNAL cancer
PATENT: WO 0157207-A 1719 09-AUG-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. .229
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 15.3%; Score 42; DB 6; Length 229;
Best Local Similarity 57.7%; Pred. No. 28;
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 124 TCTCCACCGTCCAAATTTATTTAGTTGTTGACAAATATCGACCGCTCTAAGTTCACACCGAC 183
Db 223 TTTTCTAAGGTGAGTAAATTTCAATGAATATATCTTTTCCACCATTTTACAAATTTGAAAC 164
QY 184 GCGTATAAGAGTTTCATTATATAATTTTACAAAAATAGCAATATAGCAATATATTTTCTTG 243
Db 163 GACTATACAGAAATCATATTTTAAATATATAATTAATTAATTAATTAATTAATTAATTAATTA 104
QY 244 ACTAAGCTTA 253
Db 103 AAGATGATTA 94

RESULT 15
LOCUS BC017775/c 504 bp mRNA linear PRI 29-JUN-2004
DEFINITION Homo sapiens cDNA clone IMAGE:4692106, partial cds.
ACCESSION BC017775
VERSION BC017775.1 GI:17389464
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 504)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
```

```
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, I.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalls, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 504)
Strausberg, R.
Direct Submission
Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 36 Row: m Column: 7.

FEATURES
source
1. .504
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4692106"
/tissue_type="Lung"
/clone_lib="NIH MGC_77"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

ORIGIN
Query Match 15.3%; Score 42; DB 8; Length 504;
Best Local Similarity 57.7%; Pred. No. 24;
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 124 TCTCCACCGTCCAAATTTATTTAGTTGTTGACAAATATCGACCGCTCTAAGTTCACACCGAC 183
Db 475 TTTTCTAAGGTGAGTAAATTTCAATGAATATATCTTTTCCACCATTTTACAAATTTGAAAC 416
QY 184 GCGTATAAGAGTTTCATTATATAATTTTACAAAAATAGCAATATATTTTCTTG 243
Db 415 GACTATACAGAAATCATATTTTAAATATATAATTAATTAATTAATTAATTAATTAATTAATTA 356
QY 244 ACTAAGCTTA 253
Db 355 AAGATGATTA 346

Search completed: January 12, 2006, 01:11:32
Job time : 2309 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:23:59 ; Search time 812 Seconds  
(without alignments)  
2790.406 Million cell updates/sec

Title: US-10-800-161-28  
Perfect score: 274  
Sequence: 1 tctagaataatagccgatac.....acgacgccgtaacattttc 274

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.Main:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	274	100.0	274	7	US-10-800-161-28
2	272.4	99.4	274	7	US-10-800-161-26
3	272.4	99.4	544	7	US-10-800-161-27
4	272.4	99.4	862	7	US-10-800-161-25
5	272.4	99.4	962	7	US-10-800-161-24
6	272.4	99.4	1700	8	US-10-800-161-3
7	272.4	99.4	1700	8	US-10-760-752-3
8	52.8	19.3	509	7	US-10-800-161-1
9	52.8	19.3	509	8	US-10-760-752-1
10	43.4	15.8	2469	5	US-10-027-632-102510
11	43.4	15.8	2469	5	US-10-027-632-102511
12	43.4	15.8	2469	5	US-10-027-632-112106
13	43.4	15.8	2469	6	US-10-027-632-102510
14	43.4	15.8	2469	6	US-10-027-632-102511
15	43.4	15.8	2469	6	US-10-027-632-112106
16	43	15.7	321	6	US-10-101-510-676
17	42	15.3	229	3	US-09-777-564-1719
18	42	15.3	229	5	US-10-015-219-1719
19	42	15.3	1197	6	US-10-350-923B-6
20	42	15.3	2026	6	US-10-170-385-22
21	42	15.3	4673	6	US-10-101-510-542
22	39.8	14.5	195	7	US-10-242-535A-36002
23	39.8	14.5	195	7	US-10-085-783A-36002

24	39.4	14.4	375	3	US-09-918-995-37451	Sequence 37451, A
25	38.8	14.2	6301	6	US-10-311-455-26	Sequence 26, Appl
26	38.4	14.0	794	5	US-10-027-632-147687	Sequence 147687, A
27	38.4	14.0	794	5	US-10-027-632-147688	Sequence 147688, A
28	38.4	14.0	794	5	US-10-027-632-147689	Sequence 147689, A
29	38.4	14.0	794	5	US-10-027-632-147687	Sequence 147687, A
30	38.4	14.0	794	5	US-10-027-632-147688	Sequence 147688, A
31	38.4	14.0	794	5	US-10-027-632-147689	Sequence 147689, A
32	37.8	13.8	319608	6	US-10-147-603-1	GENERAL INFORMATION
33	37.6	13.7	513	4	US-09-925-065A-534423	Sequence 534423, A
34	37.2	13.6	462	7	US-10-242-535A-7701	Sequence 7701, Ap
35	37.2	13.6	462	7	US-10-085-783A-7701	Sequence 2, Appl
36	37.2	13.6	1852	3	US-09-813-356-2	Sequence 1, Appl
37	37.2	13.6	1963	3	US-09-813-356-3	Sequence 3, Appl
38	37.2	13.6	1963	3	US-09-813-356-3	Sequence 585231, A
39	36	13.1	509	4	US-09-925-065A-585231	Sequence 47, Appl
40	36	13.1	182328	7	US-10-235-192A-47	Sequence 6, Appl
41	35.6	13.0	87394	8	US-10-810-788A-6	Sequence 69, Appl
42	35.4	12.9	140040	7	US-10-275-762-69	Sequence 2, Appl
43	35.4	12.9	367378	6	US-10-312-841-2	Sequence 10963, A
44	35.2	12.8	621	7	US-10-282-122A-10963	Sequence 2117, Ap
45	35	12.8	516	6	US-10-029-386-2117	

## ALIGNMENTS

RESULT 1  
US-10-800-161-28  
; Sequence 28, Application US/10800161  
; Publication No. US20040154051A1  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Lawton, Kay Ann  
; APPLICANT: Dietrich, Robert A  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 274  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-800-161-28

Query Match	100.0%	Score 274;	DB 7;	Length 274;
Best Local Similarity	100.0%	Pred. No. 8.1e-61;		
Matches 274;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TCTAGAATAATAGCCGATACCAATTTTCCACACATGACATTCCTTTATTCCTTTTCCAAAAGTCA	60	
Db	1	TCTAGAATAATAGCCGATACCAATTTTCCACACATGACATTCCTTTATTCCTTTTCCAAAAGTCA	60	
Qy	61	ATAAAGTGACGTCATGATACCTTTACGCTTTTAAACATCGCATGATGATGATGATGATGAT	120	
Db	61	ATAAAGTGACGTCATGATACCTTTACGCTTTTAAACATCGCATGATGATGATGATGATGAT	120	
Qy	121	CAATCTCACCGTCCTCAATTTATTTAGTTGTTGACATATCGACGCTTAAGTTCCACACC	180	
Db	121	CAATCTCACCGTCCTCAATTTATTTAGTTGTTGACATATCGACGCTTAAGTTCCACACC	180	
Qy	181	GACCGCTATAGAGTTTCATTTATTAATTTTGTAGCAAAATAAATCAGCAAAATAATTTTTC	240	
Db	181	GACCGCTATAGAGTTTCATTTATTAATTTTGTAGCAAAATAAATCAGCAAAATAATTTTTC	240	
Qy	241	TTGACTAAGCTTAAACGACGCGTTTAAACATTTTC	274	

```
Db      241 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 274
; ORGANISM: Arabidopsis thaliana
US-10-800-161-27

Query Match      99.4%; Score 272.4; DB 7; Length 544;
Best Local Similarity 99.6%; Pred. No. 2.7e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCTAGAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 60
Db      271 TCTAAAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 330
QY      61 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120
Db      331 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 390
QY      121 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 180
Db      391 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 450
QY      181 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
Db      451 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 510
QY      241 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 274
Db      511 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 544

RESULT 4
US-10-800-161-25
; Sequence 25, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089C1P1
; CURRENT APPLICATION NUMBER: US/10/800,161
; PRIOR FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 862
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-25

Query Match      99.4%; Score 272.4; DB 7; Length 862;
Best Local Similarity 99.6%; Pred. No. 3.2e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCTAGAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 60
Db      578 TCTAAAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 637
QY      61 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120
Db      638 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 697
QY      121 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 180
Db      698 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 757
QY      181 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
Db      758 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 817
QY      241 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 274

; ORGANISM: Arabidopsis thaliana
US-10-800-161-26

Query Match      99.4%; Score 272.4; DB 7; Length 274;
Best Local Similarity 99.6%; Pred. No. 2.1e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCTAGAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 60
Db      1 TCTAAAAATATAGCGGATACCAATTTTCCACACATGCGACTTCCTTTATTCCAAAGTCA 60
QY      61 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120
Db      61 ATAAAGTGTGAGCTCATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120
QY      121 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 180
Db      121 CAATCTCCACCGTCCAAATTTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 180
QY      181 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
Db      181 GACGGCTATAAGAGTTTCATTATATAATTTTAGCAAAATATAAATCAGCAAAATATTTTTC 240
QY      241 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 274
Db      241 TTGACTAAGCTTAAAGCAGCGCGTTAACATTTTC 274

RESULT 3
US-10-800-161-27
; Sequence 27, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089C1P1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 544
; TYPE: DNA
```

```

Db      818  TTGACTAAGCTTAAACGACGCCGTTAACAATTTTC 851
|||||
RESULT 5
US-10-800-161-24
; Sequence 24, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-800-161-24

Query Match      99.4%; Score 272.4; DB 7; Length 962;
Best Local Similarity 99.6%; Pred. No. 3.3e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  TCTAGAAATATAGCGGATACCAATTTTTCACACATGAGCTTCCTTTATTCCAAAGTCA 60
Db      678  TCTAAAAATATAGCGGATACCAATTTTTCACACATGAGCTTCCTTTATTCCAAAGTCA 737

Qy      61  ATAAAGTGTGACGTCATGATACCTTTACCGTTTAAAAACATCGCATGATGTCATTAGCAT 120
Db      738  ATAAAGTGTGACGTCATGATACCTTTACCGTTTAAAAACATCGCATGATGTCATTAGCAT 797

Qy      121  CAATCTCCACCGTCCAAATTTATTTAGTTGTGTGACAATATCGACCGTCTTAAGTTCACACC 180
Db      798  CAATCTCCACCGTCCAAATTTATTTAGTTGTGTGACAATATCGACCGTCTTAAGTTCACACC 857

Qy      181  GACGGCTATAAGAGTTTCATTATATAAATTTAGCAAAATATAAATCAGCAATAATTTTTC 240
Db      858  GACGGCTATAAGAGTTTCATTATATAAATTTAGCAAAATATAAATCAGCAATAATTTTTC 917

Qy      241  TTGACTAAGCTTAAACGACGCCGTTAACAATTTTC 274
Db      918  TTGACTAAGCTTAAACGACGCCGTTAACAATTTTC 951

RESULT 6
US-10-800-161-3
; Sequence 3, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1700

```

```
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (365)..(374)
; OTHER INFORMATION: TC1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (426)..(435)
; OTHER INFORMATION: TC1 motif
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (609)..(614)
; OTHER INFORMATION: MYCATR22 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (646)..(665)
; OTHER INFORMATION: CANV AS1 salicylic acid response element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (707)..(712)
; OTHER INFORMATION: PAL BOX
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (757)..(762)
; OTHER INFORMATION: HEXAMERAT 4 element
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (863)..(1228)
; OTHER INFORMATION: NI16 genomic coding region
US-10-760-752-3
```

```
Query Match 99.4%; Score 272.4; DB 8; Length 1700;
Best Local Similarity 99.6%; Pred. No. 4.1e-60;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTAGAAATATAGCGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 60
DB 578 TCTAAATAATATAGCGATACCAATTTTCCACATGGACTTCCTTTATTCCAAAGTCA 637

QY 61 ATAAAGTGAGCGTCATGATCTTAGCGCTTTAAACATCGCATGATGTCATTAGCAT 120
DB 638 ATAAAGTGAGCGTCATGATCTTAGCGCTTTAAACATCGCATGATGTCATTAGCAT 697

QY 121 CAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 180
DB 698 CAATCTCCACCGTCCCAATTTATTAGTTGTTGACAAATATCGACCGTCTAAGTTCCACACC 757

QY 181 GACGGCTATAAGGTTTCATTATTAATTTTACCAATAAATACGCAATAATATTTTTC 240
DB 758 GACGGCTATAAGGTTTCATTATTAATTTTACCAATAAATATTTAGCAATAAATATTTTTC 817

QY 241 TTGACTAAAGCTTAAACGACCGCGTTAAACATTTTC 274
DB 818 TTGACTAAAGCTTAAACGACCGCGTTAAACATTTTC 851
```

```
RESULT 8
US-10-800-161-1
; Sequence 1, Application US/10800161
; Publication No. US20040154051A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; APPLICANT: Lawton, Kay Ann
; TITLE OF INVENTION: INDUCIBLE PROMOTERS
; FILE REFERENCE: A-31089CIP1
; CURRENT APPLICATION NUMBER: US/10/800,161
```

```
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (68)..(433)
; OTHER INFORMATION: Gene product NI16
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142)..(147)
; OTHER INFORMATION: Sali site
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (344)..(349)
; OTHER INFORMATION: EcoRI site
US-10-800-161-1
```

```
Query Match 19.3%; Score 52.8; DB 7; Length 509;
Best Local Similarity 96.4%; Pred. No. 0.0015;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 219 AAAATCAGCAATAATTTTCTTCTGACTTAAGCTTAAACGACGCCGTTAACATTTTC 274
DB 1 AAAATCAGCAATAATTTTCTTCTGACTTAAGCTTAAACGACGCCGTTAACATTTTC 56
```

```
RESULT 9
US-10-760-752-1
; Sequence 1, Application US/10760752
; Publication No. US20040248303A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; TITLE OF INVENTION: Inducible Promoter Fragment
; FILE REFERENCE: A-31089DIV
; CURRENT APPLICATION NUMBER: US/10/760,752
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: 09/733685
; PRIOR FILING DATE: 2000-12-8
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (68)..(433)
; OTHER INFORMATION: gene product NI16
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142)..(147)
; OTHER INFORMATION: Sali site
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (344)..(349)
; OTHER INFORMATION: EcoRI site
US-10-760-752-1

Query Match 19.3%; Score 52.8; DB 8; Length 509;
Best Local Similarity 96.4%; Pred. No. 0.0015;
```





US-10-027-632-112106

Query Match 15.8%; Score 43.4; DB 5; Length 2469;  
Best Local Similarity 50.2%; Pred. No. 0.74;  
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 AATTTTCCACACATGGACTTCCTTTATTCACAAAGTCAATAAGTGTGACGTCAATGATA 81  
DB 301 AAAATTTATCTGAAATATATCTTCTTTATTTGAGAATTTTCACAAAATTAAGATCTCAAAAAC 360

QY 82 CTTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTA 141  
DB 361 AGCATAGTTTGTAGAGAATAAATATCAATACAGATGAATTTCTAATGTCAAAAATA 420

QY 142 TTTAGTTGTGCAATATCGACCGTCTAAGTTTCCACACCGCGGTATAAAGAGTTTCATT 201  
DB 421 ATTTCTTAAAGGATTCACAAAATGAAAAAACATAATAATCCGGCTTGAAAAACCTCATT 480

QY 202 ATAAATTTTAGCAAAATAAATCAGCAATAAT 234  
DB 481 TTATATTTTCCCAAAATAAAGACAGAGGAAT 513

## RESULT 13

US-10-027-632-102510/c  
; Sequence 102510, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 102510

; LENGTH: 2469

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-102510

Query Match 15.8%; Score 43.4; DB 6; Length 2469;  
Best Local Similarity 50.2%; Pred. No. 0.74;  
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 AATTTTCCACACATGGACTTCCTTTATTCACAAAGTCAATAAGTGTGACGTCAATGATA 81  
DB 2169 AAAATTTATCTGAAATATATCTTCTTTATTTGAGAATTTTCACAAAATTAAGATCTCAAAAAC 2110

QY 82 CTTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTA 141  
DB 2109 AGCATAGTTTGTAGAGAATAAATATCAATACAGATGAATTTCTAATGTCAAAAATA 2050

QY 142 TTTAGTTGTGCAATATCGACCGTCTAAGTTTCCACACCGCGGTATAAAGAGTTTCATT 201  
DB 2049 ATTTCTTAAAGGATTCACAAAATGAAAAAACATAATAATCCGGCTTGAAAAACCTCATT 1990

QY 202 ATAAATTTTAGCAAAATAAATCAGCAATAAT 234

DB 1989 TTATATTTGCCAAAATAAAGACAGAGGAAT 1957

## RESULT 14

US-10-027-632-102511/c  
; Sequence 102511, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 102511

; LENGTH: 2469

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-102511

Query Match 15.8%; Score 43.4; DB 6; Length 2469;  
Best Local Similarity 50.2%; Pred. No. 0.74;  
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 AATTTTCCACACATGGACTTCCTTTATTCACAAAGTCAATAAGTGTGACGTCAATGATA 81  
DB 2169 AAAATTTATCTGAAATATATCTTCTTTATTTGAGAATTTTCACAAAATTAAGATCTCAAAAAC 2110

QY 82 CTTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTA 141  
DB 2109 AGCATAGTTTGTAGAGAATAAATATCAATACAGATGAATTTCTAATGTCAAAAATA 2050

QY 142 TTTAGTTGTGCAATATCGACCGTCTAAGTTTCCACACCGCGGTATAAAGAGTTTCATT 201  
DB 2049 ATTTCTTAAAGGATTCACAAAATGAAAAAACATAATAATCCGGCTTGAAAAACCTCATT 1990

QY 202 ATAAATTTTAGCAAAATAAATCAGCAATAAT 234  
DB 1989 TTATATTTGCCAAAATAAAGACAGAGGAAT 1957

## RESULT 15

US-10-027-632-112106  
; Sequence 112106, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112106
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-112106

Query Match      15.8%; Score 43.4; DB 6; Length 2469;
Best Local Similarity 50.2%; Pred. No. 0.74;
Matches 107; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY      22 AATTTTCCACATCGACTTCCTTTATTCCAAAAGTCAATAAAGTGTGACGTCATGATA 81
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 AAAATTATCTGAAATATATCTTCTTATTGAGAATTTACAAAATTTAAGATCTCAAAAAC 360

QY      82 CTTACGCTTTAAACATCGCATGATGTCATTCATTCATCAATCTCCACCGTCCAATTTA 141
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
361 AGCATAGTTTGTAGTAAAGAAATAAATATCATTTACAGAATGAAATCTTAATGTCAAAAATA 420

QY      142 TTTAGTTGTTGACAATATCGACCGTCTAAGTTCACACCGCGCTATAAGAGTTTTCATT 201
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 ATTCTTAAAGGATTCACCAAAATGAAACATAATAATCCGCTTGAAAAACCTCATT 480

QY      202 ATAAATTTTAGCAAAATAAATCAGCAAAATAAT 234
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 TTATATTGCCAAAATAAGAACAGAGGAAT 513
```

Search completed: January 12, 2006, 02:31:42  
Job time : 817 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:33:06 ; Search time 395 Seconds  
(without alignments)  
561.421 Million cell updates/sec

Title: US-10-800-161-28  
Perfect score: 274  
Sequence: 1 tctagaataatagcgatac.....acgagccgtaacattttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6038814 seqs, 404674181 residues

Total number of hits satisfying chosen parameters: 12077628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:
- 5: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:
- 7: /cgn2\_6/ptodata/2/pubpna/US12\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/2/pubpna/US13\_NEW\_PUB.seq:
- 9: /cgn2\_6/ptodata/2/pubpna/US14\_NEW\_PUB.seq:
- 10: /cgn2\_6/ptodata/2/pubpna/US15\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	39	14.2	197096	7 US-11-121-086-107	Sequence 107, App
c 2	37.8	13.8	319608	7 US-11-145-703-1	Sequence 1, Appl
c 3	32.6	11.9	1082144	7 US-11-117-187-211	Sequence 211, App
c 4	32	11.7	1765	6 US-10-750-185-59119	Sequence 59119, A
c 5	32	11.7	1765	6 US-10-750-623-59119	Sequence 59119, A
c 6	31.8	11.6	100000	7 US-11-124-368A-2918	Sequence 2918, Ap
c 7	31.4	11.5	3857	7 US-11-136-527-3203	Sequence 3203, Ap
c 8	31.2	11.4	5152	6 US-10-240-708-73	Sequence 73, Appl
c 9	31	11.3	1140	6 US-10-750-185-39222	Sequence 39222, A
c 10	31	11.3	1272	6 US-10-750-185-39222	Sequence 39222, A
c 11	31	11.3	1272	6 US-10-750-185-37551	Sequence 37551, A
c 12	31	11.3	1400	7 US-11-136-527-2599	Sequence 2599, Ap
c 13	31	11.3	2217	7 US-11-136-527-2599	Sequence 2599, Ap
c 14	31	11.2	4627	6 US-10-775-169-206	Sequence 206, App
c 15	30.6	11.2	4627	6 US-11-186-284-52	Sequence 52, Appl
c 16	30.6	11.2	4627	7 US-11-121-086-41	Sequence 41, Appl
c 17	30.6	11.2	181172	7 US-11-121-086-41	Sequence 4185, Ap
c 18	30.4	11.1	600	7 US-11-136-527-4185	Sequence 7677, Ap
c 19	30.4	11.1	1538	6 US-10-750-185-26322	Sequence 26322, A
c 20	30.4	11.1	1538	6 US-10-750-185-26322	Sequence 26322, A
c 21	30.4	11.1	1538	6 US-10-750-623-26322	Sequence 89, Appl
c 22	30.4	11.1	1791	7 US-11-136-527-89	Sequence 89, Appl
c 23	30.4	11.1	2043	7 US-11-136-527-3581	Sequence 3581, Ap

c 24	30.4	11.1	2990	6 US-10-793-626-3899	Sequence 3899, Ap
c 25	30.4	11.1	5037	6 US-10-415-198A-13	Sequence 13, Appl
c 26	30.2	11.0	729	6 US-10-750-185-31590	Sequence 31590, A
c 27	30.2	11.0	729	6 US-10-750-623-31590	Sequence 31590, A
c 28	30.2	11.0	3629	6 US-10-750-185-56349	Sequence 56349, A
c 29	30.2	11.0	3629	6 US-10-750-623-56349	Sequence 56349, A
c 30	30.2	11.0	5535	6 US-10-240-708-17	Sequence 17, Appl
c 31	30.2	11.0	142605	7 US-11-121-086-64	Sequence 64, Appl
c 32	30.2	11.0	186442	7 US-11-121-086-104	Sequence 104, App
c 33	30.2	11.0	235033	7 US-11-157-389-1	Sequence 1, Appl
c 34	30.2	11.0	237326	7 US-11-157-389-2	Sequence 2, Appl
c 35	30	10.9	959	6 US-10-750-185-36299	Sequence 36299, A
c 36	30	10.9	959	6 US-10-750-623-36299	Sequence 36299, A
c 37	30	10.9	1220	6 US-10-750-185-29212	Sequence 29212, A
c 38	30	10.9	1220	6 US-10-750-623-29212	Sequence 29212, A
c 39	30	10.9	178877	7 US-11-121-086-17	Sequence 17, Appl
c 40	29.8	10.9	752	6 US-10-750-185-36704	Sequence 36704, A
c 41	29.8	10.9	752	6 US-10-750-623-36704	Sequence 36704, A
c 42	29.8	10.9	1210	6 US-10-750-185-37833	Sequence 37833, A
c 43	29.8	10.9	1210	6 US-10-750-623-37833	Sequence 37833, A
c 44	29.8	10.9	3032	6 US-10-793-626-4227	Sequence 4227, Ap
c 45	29.8	10.9	3099	6 US-10-392-234A-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1

US-11-121-086-107/c  
; Sequence 107, Application US/11121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; APPLICANT: NIELSEN, KIRSTEN V.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; PRIOR FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 107  
; LENGTH: 197096  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-11-121-086-107

Query Match	14.2%	Score 39;	DB 7;	Length 197096;
Best Local Similarity	51.4%	Pred. No. 1.2;		
Matches	90;	Conservative	0;	Mismatches 85; Indels 0; Gaps 0;
Qy	69	TGAGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCC	128	
Db	27553	TGTATTTCATGACAGTTTCTATATATAATGAACATTAGGGTTCATATATTTCTTTT	27494	
Qy	129	ACCGTCCAAATTTTATTTAGTTGTGACAATATCGACCGCTCTAAAGTTCCACACCGGCTA	188	
Db	27493	GGCCTCTCATTGTGCTATTCATTCATGATCATCTTTCCATCAGCACACACAGCTG	27434	
Qy	189	TAAGAGTTTCATTATAAATTTAGCAAAATAAATACGAAATAATTTTTCTTG	243	
Db	27433	TAAATCATTTGTTAATAGATGATACCAATATGATTATTAAGTCTCTTTCTTTG	27379	

RESULT 2

US-11-145-703-1  
; Sequence 1, Application US/11145703  
; Publication No. US20050260667A1  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Chumakov, Ilya

```
/ APPLICANT: Bougueleret, Lydie
/ APPLICANT: Bihain, Bernard
/ APPLICANT: Essioux, Laurent
/ TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
/ FILE REFERENCE: 53 US16.DIV
/ CURRENT APPLICATION NUMBER: US/11/145,703
/ CURRENT FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: US/10/147,603
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: 09/539,333
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: US 60/126,903
/ PRIOR FILING DATE: 1999-03-30
/ PRIOR APPLICATION NUMBER: US 60/131,971
/ PRIOR FILING DATE: 1999-04-30
/ PRIOR APPLICATION NUMBER: US 60/132,065
/ PRIOR FILING DATE: 1999-04-30
/ PRIOR APPLICATION NUMBER: US 60/143,928
/ PRIOR FILING DATE: 1999-07-14
/ PRIOR APPLICATION NUMBER: US 60/145,915
/ PRIOR FILING DATE: 1999-07-27
/ PRIOR APPLICATION NUMBER: US 60/146,453
/ PRIOR FILING DATE: 1999-07-29
/ PRIOR APPLICATION NUMBER: US 60/146,452
/ PRIOR FILING DATE: 1999-07-29
/ PRIOR APPLICATION NUMBER: US 60/162,288
/ PRIOR FILING DATE: 1999-10-28
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 234
/ SOFTWARE: Patent.pm
/ SEQ ID NO 1
/ LENGTH: 319608
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 31..1107
/ OTHER INFORMATION: 5'regulatory region g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 1108..1289
/ OTHER INFORMATION: exon A g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 14877..14920
/ OTHER INFORMATION: exon B g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 18778..18862
/ OTHER INFORMATION: exon Bbis g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 25593..25740
/ OTHER INFORMATION: exon C g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 29388..29502
/ OTHER INFORMATION: exon D g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 29967..30282
/ OTHER INFORMATION: exon E g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 64666..64812
/ OTHER INFORMATION: exon F g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 65505..65853
/ OTHER INFORMATION: exon G g35018 gene
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 65854..67854
/ OTHER INFORMATION: exon Nbis complement g34872 gene

/ OTHER INFORMATION: 3'regulatory region g35018 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 94124..94964
/ OTHER INFORMATION: exon g35017
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 201188..201234
/ OTHER INFORMATION: exon S g35030 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 214676..214793
/ OTHER INFORMATION: exon T g35030 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 215702..215746
/ OTHER INFORMATION: exon U g35030 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 216836..216915
/ OTHER INFORMATION: exon V g35030 gene
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 213818..215818
/ OTHER INFORMATION: 3'regulatory region g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 215819..215941
/ OTHER INFORMATION: exon R complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 215819..215975
/ OTHER INFORMATION: exon Rbis complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 216661..216952
/ OTHER INFORMATION: exon Qbis complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 217027..217061
/ OTHER INFORMATION: exon Q1 complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 229647..229742
/ OTHER INFORMATION: exon X complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 230408..230721
/ OTHER INFORMATION: exon P complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 231272..231412
/ OTHER INFORMATION: exon Obis complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 231787..231880
/ OTHER INFORMATION: exon O2 complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 231870..231879
/ OTHER INFORMATION: exon O1 complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 234174..234321
/ OTHER INFORMATION: exon O complement g34872 gene
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 237406..237428
/ OTHER INFORMATION: exon Nbis complement g34872 gene
```

```

FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon MS2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240994
OTHER INFORMATION: exon M692 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240993
OTHER INFORMATION: exon MS1 complement g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 241686..243685
OTHER INFORMATION: 5' regulatory region g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 290652..292652
OTHER INFORMATION: 3' regulatory region g34665 gene

```

Query Match	13.8%	Score 37.8	DB 7	Length 319608
Best Local Similarity	52.9%	Pred. No. 2.9		
Matches	81	Conservative 0	Mismatches 72	Indels 0
Gaps	0			
Qy	38	GACTTCCTTTATTCCAAAAGTCAATAAAGTGTGACGTGATGATCTTACGCTTTAAAAACA	97	
Db	57795	GAACTTGATCTTCCCAACGTGTCTACATGGTGACCTTCTTAATTCACCCATATATACAG	57854	
Qy	98	TCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTTAGTTGTTCACAAT	157	
Db	57855	TTGCTGTGTAATGCTTTTATCTTCAATGTCAGCGGTCCAAAGAGGTAAAAAGAGAAAAAC	57914	
Qy	158	ATCGACCGTCTAAGTTCCACACCGACGGCTATA	190	
Db	57915	AAAGGGAGAAAAAATGAGGCACACACCTCTTTTA	57947	

RESULT 3  
US-11-117-187-211/c  
; Sequence 211, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY

```

; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
;
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1993-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-11-117-187-211

```

	Query Match	11.9%	Score 32.6	DB 7	Length 1082144
	Best Local Similarity	50.3%	Pred. No. le+02		
	Matches 80	Conservative 0	Mismatches 79	Indels 0	Gaps 0
Qy	81	ACTTACGCTTTAAACAATCGCATGATGTCATTAGCATCAATCTCCACGCGTCCAAATTT	140		
Db	345257	ATTGGTGCTTAAACAATTTCTTAACTTTTATGATCCTATTCAATCTAAACCGTTCAAATAA	345198		
Qy	141	ATTTAGTGTGTTGACAATATCGACCGTCTTAAGTTCACACCGCGCTATAGAAGTTTCAT	200		
Db	345197	AAATTATTTCCCTTCAATATATAAACCGTTCACAAAAATAAGGAATCTATGTCGTTTTTGAA	345138		
Qy	201	TATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTTT	239		
b	345137	ATCAAAATATATACATTTTCATCTCCCTAAAAACACTATAT	345099		

## RESULT 4

```

US-10-750-185-59119
; Sequence 59119, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 59119
; LENGTH: 1765
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-59119

```

```

Query Match      11.7%; Score 32; DB 6; Length 1765;
Best Local Similarity 56.7%; Pred. No. 20;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 145 AGTTGTTGACAATATCGACCGTCTTAAGTTCACACCGCCTATAGAGTTTCATTATA 204
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 936 AGTTGAAAACCTCTGCCACACAGAAATCTGCACACAGAGGTTTACAGCAGTTTAGTCATA 995

QY 205 AATTTTAGCAAAATAAAATACGAAATAATTTTTTCTTGACTAA 248
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 996 ATTGCAAAATTTGGAAGCCACCAAGATATTTCTTTGTTAGTAA 1039

RESULT 5

```







```
; TYPE: DNA
; ORGANISM: Bovine 198668063412
US-10-750-185-37551

Query Match
Best Local Similarity 11.3%; Score 31; DB 6; Length 1272;
Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 55 AAGTCAATAAGTGTGACGTATGATCTTAAAGCTTTAAACATCGCATGATGATGTCAT 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1177 AACAAATATCAAGATAGGTGATGAGGTACCCAGATTACAGAGCCTTACTATTCTAA 1118

QY 115 TAGCATCAATCTCCACCGTCCCAATTTATTAGTGTGACATATCGACCGTCTAAGTTC 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1117 AAGTGATATACCAAAATAACATCTATCTGTTGTTTAAATGTAGAACCTCAAGGTCC 1058

QY 175 CACACCGCGGTCTATAAGAGTTTCTATTATA 205
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1057 CACCCTCACATATATTAACCTTCATTTTAA 1027

RESULT 12
US-10-750-623-37551/c
; Sequence 37551, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37551
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Bovine 198668063412
US-10-750-623-37551

Query Match
Best Local Similarity 11.3%; Score 31; DB 6; Length 1272;
Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 55 AAGTCAATAAGTGTGACGTATGATCTTAAAGCTTTAAACATCGCATGATGATGTCAT 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1177 AACAAATATCAAGATAGGTGATGAGGTACCCAGATTACAGAGCCTTACTATTCTAA 1118

QY 115 TAGCATCAATCTCCACCGTCCCAATTTATTAGTGTGACATATCGACCGTCTAAGTTC 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1117 AAGTGATATACCAAAATAACATCTATCTGTTGTTTAAATGTAGAACCTCAAGGTCC 1058

QY 175 CACACCGCGGTCTATAAGAGTTTCTATTATA 205
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1057 CACCCTCACATATATTAACCTTCATTTTAA 1027

RESULT 13
US-11-136-527-6695
; Sequence 6695, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
```

```
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6695
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6695

Query Match
Best Local Similarity 11.3%; Score 31; DB 7; Length 1400;
Matches 49; Conservative 15; Mismatches 45; Indels 0; Gaps 0;

QY 138 TTTATTAGTTGTTGACAAATATCGACCGTCTAAAGTTCACACCGCGGTATAAGAGTTT 197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 TTTCTGTATTACTTAAATTCATGCTCTTCAATATACAAAAATAAGTAAGATGTTCT 248

QY 198 CATTATAAATTTAGCAAAATAAATCAGCAAAATAATTTTCTTGACT 246
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 CCTAAWAWWWWWWWMAAAARRWWMAWAWMTAGTTTCTTTTACT 297

RESULT 14
US-11-136-527-2599
; Sequence 2599, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2599
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2599

Query Match
Best Local Similarity 11.3%; Score 31; DB 7; Length 2217;
Matches 49; Conservative 15; Mismatches 45; Indels 0; Gaps 0;

QY 138 TTTATTAGTTGTTGACAAATATCGACCGTCTAAAGTTCACACCGCGGTATAAGAGTTT 197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1006 TTTCTGTATTACTTAAATTCATGCTCTTCAATATACAAAAATAAGTAAGATGTTCT 1065

QY 198 CATTATAAATTTAGCAAAATAAATCAGCAAAATAATTTTCTTGACT 246
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1066 CCTAAWAWWWWWWWMAAAARRWWMAWAWMTAGTTTCTTTTACT 1114

RESULT 15
US-10-775-169-206
; Sequence 206, Application US/10775169
; Publication No. US20050287532A9
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dornier, Andrew
; APPLICANT: Trepicchio, William
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
; FILE REFERENCE: AM101080 (031896-013000)
; CURRENT APPLICATION NUMBER: US/10/775,169
; CURRENT FILING DATE: 2004-02-11
```

NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 206  
; LENGTH: 4627  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-206

Query Match 11.2%; Score 30.6; DB 6; Length 4627;  
Best Local Similarity 47.2%; Pred. No. 64;  
Matches 93; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy	45	TTTATTCCRAAAGTCAATAAAGTGTGACGTCATGATACTTTACGCTTTTAAACATCGCATG	104
Db	4361	TTTTATTAAAGATGCTATGGAACATAAAGTTGTATGCCATGCAATTTAAAGTAACTTATT	4420
Qy	105	ATGATGTCAATTAGCATCAATCTCCACCGTCCAAATTTATTAGTTGTGTGACAATATCGACC	164
Db	4421	TGACTATGAATATTATCGGATTACTGAATTGTATCAATTTGTTGTGTTCAATATCAGCT	4480
Qy	165	GTCTAAGTTCACACCGACCGCTATAAGAGTTTCATTATATAATTTAGCAAAATAAATC	224
Db	4481	TTGATRAATTGTGTACCTTAAAGATATTGAGGAGAGAAAATAGATAAATTACAAAGATATTATT	4540
Qy	225	AGCAATAAATTTTCT	241
Db	4541	AAATTTTATTATTATT	4557

Search completed: January 12, 2006, 02:38:32  
Job time : 401 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:16:10 ; Search time 148 Seconds  
(without alignments)  
3290.891 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274

Sequence: 1 tctagaatatccgctac.....acgaacgcgttaacattttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq.\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq.\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272.4	99.4	1700	3	US-09-733-685-3
2	52.8	19.3	509	3	US-09-733-685-1
3	39	14.2	601	3	US-09-949-016-83537
4	39	14.2	49440	3	US-09-949-016-14150
5	37.8	13.8	319608	3	US-09-539-333D-1
6	37.8	13.8	319608	3	US-09-679-409-1
7	37.2	13.6	1852	3	US-09-813-356-2
8	37.2	13.6	1963	3	US-09-813-356-1
9	37.2	13.6	1963	3	US-09-813-356-3
10	35.4	12.9	30002	3	US-09-949-016-15866
11	34.6	12.6	1141	3	US-09-806-708B-22
12	34.4	12.6	227390	3	US-09-949-016-12201
13	34.4	12.6	227391	3	US-09-949-016-13365
14	34.2	12.5	601	3	US-09-949-016-112718
15	34.2	12.5	601	3	US-09-949-016-112802
16	34.2	12.5	601	3	US-09-949-016-112886
17	34.2	12.5	601	3	US-09-949-016-112962
18	34.2	12.5	601	3	US-09-949-016-113048
19	34.2	12.5	127280	3	US-09-949-016-14857
20	34.2	12.5	132266	3	US-09-949-016-14860
21	34.2	12.5	150833	3	US-09-949-016-14859
22	34.2	12.5	159963	3	US-09-949-016-14858
23	34.2	12.5	171130	3	US-09-949-016-14861
24	34	12.4	2275	2	US-08-743-637B-2

25	34	12.4	2275	3	US-08-526-840B-2	Sequence 2, Appli
26	33.6	12.3	474	3	US-09-621-976-18033	Sequence 18033, A
27	33.6	12.3	3244	3	US-09-949-016-2708	Sequence 2708, Ap
28	33.6	12.3	3953	3	US-09-220-132-46	Sequence 46, Appl
29	33.6	12.3	3953	3	US-09-949-016-263	Sequence 263, App
30	33.6	12.3	12797	3	US-09-949-016-15274	Sequence 15274, A
31	33.4	12.2	832	3	US-09-621-976-2813	Sequence 2813, Ap
32	33.4	12.2	27363	3	US-09-949-016-17535	Sequence 17535, A
33	33.4	12.2	67755	3	US-09-949-016-13703	Sequence 13703, A
34	33.2	12.1	466	2	US-08-743-637B-29	Sequence 29, Appl
35	33.2	12.1	466	3	US-08-526-840B-29	Sequence 29, Appl
36	33.2	12.1	600	3	US-09-540-236-1274	Sequence 1274, Ap
37	33.2	12.1	92407	3	US-09-596-002-36	Sequence 36, Appl
38	33.2	12.1	139049	3	US-09-949-016-17030	Sequence 17030, A
39	33.2	12.1	139936	3	US-09-949-016-11782	Sequence 11782, A
40	33.2	12.1	139952	3	US-09-949-016-13280	Sequence 13280, A
41	33	12.0	480	3	US-09-621-976-13994	Sequence 13994, A
42	33	12.0	10320	3	US-09-949-016-11778	Sequence 11778, A
43	32.8	12.0	198	3	US-09-513-999C-19031	Sequence 19031, A
44	32.8	12.0	893	3	US-09-270-767-15009	Sequence 15009, A
45	32.8	12.0	1664976	3	US-08-916-421B-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-733-685-3  
; Sequence 3, Application US/09733685  
; Patent No. 6706952  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; TITLE OF INVENTION: GENES ENCODING PROTEINS INVOLVED IN THE REGULATION OF  
; FILE REFERENCE: A-31089A  
; CURRENT APPLICATION NUMBER: US/09/733,685  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1700  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (365)..(374)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (426)..(435)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (609)..(614)  
; OTHER INFORMATION: MYCATR22 element  
; NAME/KEY: misc feature  
; LOCATION: (646)..(665)  
; OTHER INFORMATION: CAMV ASI salicylic acid response element  
; NAME/KEY: misc feature  
; LOCATION: (707)..(712)  
; OTHER INFORMATION: PAL BOX  
; NAME/KEY: misc feature  
; LOCATION: (757)..(762)  
; OTHER INFORMATION: HEXAMERAT 4 element  
; NAME/KEY: misc feature  
; LOCATION: (863)..(1228)  
; OTHER INFORMATION: NIL6 genomic coding region  
US-09-733-685-3

Query Match 99.4%; Score 272.4; DB 3; Length 1700;



```
QY      189 TAAGATTCATTATAATTTTACGAAAAATAAAATCAGCAATAATTTTCTTG 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      24627 TAACATCATGTTAATAGATGACCAATAATGATTTAATAGTCCTTCTTCTTG 24681

RESULT 5
US-09-539-333D-1
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:

; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
```





LOCATION: 244353..244561	OTHER INFORMATION: 8-293-130	: polymorphic base A or G
OTHER INFORMATION: exon Y	NAME/KEY: allele	
LOCATION: 246273..247802	OTHER INFORMATION: 8-292-198	: polymorphic base A or G
OTHER INFORMATION: exon Z	NAME/KEY: allele	
LOCATION: 247803..249803	OTHER INFORMATION: 8-251-322	: polymorphic base A or G
OTHER INFORMATION: 3'regulatory region	NAME/KEY: allele	
LOCATION: 8316	OTHER INFORMATION: 8-289-322	: polymorphic base A or G
OTHER INFORMATION: 99-27943-150	NAME/KEY: allele	
LOCATION: 21672	OTHER INFORMATION: 8-287-249	: polymorphic base C or T
OTHER INFORMATION: 99-27935-193	NAME/KEY: allele	
LOCATION: 65485	OTHER INFORMATION: 8-287-86	: polymorphic base A or T
OTHER INFORMATION: 8-128-33	NAME/KEY: allele	
LOCATION: 95396	OTHER INFORMATION: 8-285-319	: polymorphic base A or G
OTHER INFORMATION: 99-31960-363	NAME/KEY: allele	
LOCATION: 107281	OTHER INFORMATION: 8-283-278	: polymorphic base G or C
OTHER INFORMATION: 99-24656-260	NAME/KEY: allele	
LOCATION: 160640	OTHER INFORMATION: 8-283-176	: polymorphic base A or G
OTHER INFORMATION: 99-24639-163	NAME/KEY: allele	
LOCATION: 160876	OTHER INFORMATION: 8-283-153	: polymorphic base G or C
OTHER INFORMATION: 99-24634-108	NAME/KEY: allele	
LOCATION: 168974	OTHER INFORMATION: 8-283-56	: polymorphic base C or T
OTHER INFORMATION: 99-7652-162	NAME/KEY: allele	
LOCATION: 170810	OTHER INFORMATION: 8-282-345	: polymorphic base G or C
OTHER INFORMATION: 99-16100-147	NAME/KEY: allele	
LOCATION: 173358	OTHER INFORMATION: 8-282-260	: polymorphic base G or T
OTHER INFORMATION: 99-5862-167	NAME/KEY: allele	
LOCATION: 189957	OTHER INFORMATION: 8-282-245	: polymorphic base A or C
OTHER INFORMATION: 99-5919-215	NAME/KEY: allele	
LOCATION: 197163	OTHER INFORMATION: 8-282-174	: variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
OTHER INFORMATION: 99-24658-410	NAME/KEY: allele	
LOCATION: 200778	OTHER INFORMATION: 8-282-92	: polymorphic base A or T
OTHER INFORMATION: 8-303-235	NAME/KEY: allele	
LOCATION: 202651	OTHER INFORMATION: 8-281-367	: polymorphic base A or G
OTHER INFORMATION: 8-300-221	NAME/KEY: allele	
LOCATION: 202679	OTHER INFORMATION: 8-281-299	: polymorphic base A or G
OTHER INFORMATION: 8-300-193	NAME/KEY: allele	
LOCATION: 203378	OTHER INFORMATION: 8-281-248	: polymorphic base G or C
OTHER INFORMATION: 8-299-128	NAME/KEY: allele	
LOCATION: 204138	OTHER INFORMATION: 8-279-197	: polymorphic base A or C
OTHER INFORMATION: 8-296-213	NAME/KEY: allele	
LOCATION: 204605	OTHER INFORMATION: 8-278-289	: polymorphic base C or T
OTHER INFORMATION: 8-252-190	NAME/KEY: allele	
LOCATION: 204934	OTHER INFORMATION: 8-252-190	: polymorphic base C or T
OTHER INFORMATION: 99-24644-194	NAME/KEY: allele	
LOCATION: 205206	OTHER INFORMATION: 99-24644-194	: polymorphic base A or G
OTHER INFORMATION: 8-295-248	NAME/KEY: allele	
LOCATION: 205329	OTHER INFORMATION: 8-295-248	: polymorphic base A or C
OTHER INFORMATION: 8-295-125	NAME/KEY: allele	
LOCATION: 206064	OTHER INFORMATION: 8-295-125	: polymorphic base C or T

Query Match	13.8%	Score 37.8;	DB 3;	Length 319608;
Best Local Similarity	52.9%	Pred. No. 1.1;		
Matches 3,81;	Conservative	0;	Mismatches 72;	Indels 0;
Gaps	0;			
Qy	38	GACTTCCTTTATCCAAAGTCAATAAGTGACGTCATGATACCTTACCGCTTTAAACA	97	
Db	57795	GAACCTGTATCTCCCACTGTGTCTACATGTCGACTCTCTAATTTCCACCATATATACAG	57854	
Qy	98	TCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTGCAAT	157	

Db 57855 TTGCTGTGAATGCTCTTTTATCTTCCTCAATGTCAGCGGTCCAAAAGAGTAGTAATAAAGAGAAAAAC 57911

QY 158 ATCAGCCGGCTTAAGTTTCCACACCGACGGCGTATA 190  
| | | | | | | | | | | | | | | | | | | |  
Db 57915 AAAGGGAGAAAAAATGAGGCACCACCTCTTTA 57947

RESULT 7  
US-09-813-356--2  
; Sequence 2, Application US/09813356  
; Patent No. 6730826  
; GENERAL INFORMATION:  
; APPLICANT: George, Wagner  
; APPLICANT: Susheng, Gan  
; APPLICANT: Erming, Wang  
; APPLICANT: Rui, Wang  
; TITLE OF INVENTION: Trichome Specific Regulatory Sequence  
; FILE REFERENCE: 50229-260  
; CURRENT APPLICATION NUMBER: US/09/813,356  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,088  
; PRIOR FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1852  
; TYPE: DNA  
; ORGANISM: Nicotiana tabacum  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1852)  
; NAME/KEY: CAAT signal  
; LOCATION: (1669)..(1672)  
; NAME/KEY: GC signal  
; LOCATION: (1806)..(1807)

US-09-813-356--2

Query Match 13.6%; Score 37.2; DB 3; Length 1852;  
Best Local Similarity 47.8%; Pred. No. 0.46;  
Matches 108; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 16 GATACCAAATTTTCCACACATGGACTTCCTTTATTCCAAAAGTCAATAAAAGTGTCAGCTC 75  
Db 730 GAAACAATAATTTTGTAGTCAATAAATAATTTATCTCTTTATTAAAGAAAATAAAAAATAG 789

QY 76 ATGATCTACTACGCTTTAAACATCGCATGATGATGTCATGCATCAATCAATCTCCACGCTC 135  
Db 790 TTAGATAAATAGTTTACTACTATTTGTCATGAAAATATCAATAGATACAAATTTTAAAGTGAC 849

QY 136 AATTTATTTAGTTGTTGACAAATATCGACCGTCTTAAGTTCACACCGACGGCTATAAGAGT 195  
Db 850 TATAAATTTACGAGTTTACTATACITTTAGTCGTACAGTTTGCAATAATAGTATTTTAACC 909

QY 196 TTCATTATAAATTTTAGCAAAATAAATAATCAGCAATAATTTTTTCT 241  
Db 910 ACAATTAGTTTATATGTACAAATAAACATAAGTGAATAACTTTTTT 955

RESULT 8  
US-09-813-356-1  
; Sequence 1, Application US/09813356  
; Patent No. 6730826  
; GENERAL INFORMATION:  
; APPLICANT: George, Wagner  
; APPLICANT: Susheng, Gan  
; APPLICANT: Erming, Wang  
; APPLICANT: Rui, Wang  
; TITLE OF INVENTION: Trichome Specific Regulatory Sequence  
; FILE REFERENCE: 50229-260  
; CURRENT APPLICATION NUMBER: US/09/813,356  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,088  
; PRIOR FILING DATE: 2000-03-22



```
; Sequence 13365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13365
; LENGTH: 227391
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(227391)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13365

Query Match      12.6%; Score 34.4; DB 3; Length 227391;
Best Local Similarity 52.0%; Pred. No. 9.3;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 83 TTACGCTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACGGTCCAAATTAT 142
Db 161269 TTTCAGAAAAAATAAATGTTGATCAATTACTATTACTTGCATATTCCTTTATCT 161210

QY 143 TTAGTTGTTGACAATATCGACGCTTAAGTTCACACCGACGCTAAGAGTTTCATTA 202
Db 161209 AATTTTGCTCTCCAAATGTCCTTGAGAACCCACAGGAGGAAGACTCTCTTTT 161150

QY 203 TAAATTTTAGCAAAATAAATCAGCAAA 230
Db 161149 TTCCTTAAGTTACATCCATCAGCAAA 161122

RESULT 14
US-09-949-016-112718/c
; Sequence 112718, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112718
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-112718

Query Match      12.5%; Score 34.2; DB 3; Length 601;
Best Local Similarity 48.2%; Pred. No. 2.5;
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 12 AGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCAATAAAGTGTA 71
Db 210 AGGCAACCACAGTTTGTCTTTTCTGAATATTTTATTTAAAGGAATTTATAAATAGGTA 151

QY 72 CGTCATGATACCTTACGCTTTTAAACATCGCATGATGTCATTAGCATCAATCTCCACC 131
Db 150 GTCCTTTTATATGCTCTTTTATATGACATAGCATATATGTCAAACTCATCTTTTGACCACACC 91

QY 132 GTCCAAATTTATTAGTTGTTGACAATATCGACCGCTTAAGTTCACACCGACGGCTATAA 191
Db 90 CTCAAATATGTTATGTCCTCCAAACCTTCCTTACCTGTAAGTGGCACTTCTATTCCAGCCA 31

QY 192 GAGTTTCATTATAAATTTT 210
Db 30 GAATTTTCAGAAAGTAATTTT 12

Search completed: January 12, 2006, 02:17:53
Job time : 155 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2006, 00:02:06 ; Search time 3802 Seconds

(without alignments)  
3371.820 Million cell updates/sec

Title: US-10-800-161-28

Perfect score: 274

Sequence: 1 tctagaatatagccgatac.....acgacgcgttaacatttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222	81.0	228	9	BZ352867 SALK_0827
2	152	55.5	152	9	BZ770600 SALK_1435
3	106.2	38.8	237	9	BH814587 SALK_0666
4	101.2	36.9	147	9	BZ352865 SALK_0827
5	63.2	23.1	691	9	BZ019485 oeg72q05
6	49	17.9	397	3	BH863458 BP863458
7	45.4	16.6	465	9	B52379 CIT-HSP-200
8	43.6	15.9	609	1	AI990512 ws40e09.x
9	43	15.7	321	8	R88734 YP93b07.sl
10	42.2	15.4	1204	10	CNS016E2
11	42	15.3	135	1	AI540689 tn72ali.x
12	42	15.3	289	1	AI873432 wf82f11.x
13	42	15.3	360	1	AI290306 qm02a04.x
14	42	15.3	418	1	AI620654 tu85c06.x
15	42	15.3	421	1	AI261719 qz31b01.x
16	42	15.3	448	1	AI357111 qx16g02.x
17	42	15.3	452	1	AA903013 ok45e06.s
18	42	15.3	462	1	AW014993 UI-H-B10p
19	42	15.3	477	1	AI151151 qc87f07.x
20	42	15.3	498	1	AW512290 xx72h07.x
21	42	15.3	553	5	BU689059 UI-CF-EC1
22	42	15.3	558	1	AW339459 xz91b03.x

23	42	15.3	610	3	BM995524
24	42	15.3	657	3	BQ015514
25	42	15.3	684	3	BQ003373
26	42	15.3	727	3	BQ006479
27	42	15.3	750	6	CA447735
28	42	15.3	2026	4	AF116653
29	41.6	15.2	283	3	BP423473
30	41.6	15.2	412	1	AI494002
31	41.6	15.2	418	1	AI708269
32	41.6	15.2	442	2	BE219714
33	41.6	15.2	553	1	AI797039
34	41.4	15.1	410	8	W96222
35	41.4	15.1	471	3	BM144644
36	41.4	15.1	473	3	BM150022
37	41.4	15.1	475	1	AA700613
38	41.4	15.1	481	1	AW051787
39	41.4	15.1	491	1	AI708670
40	41.4	15.1	534	2	BE856347
41	41.4	15.1	548	2	BE501745
42	40.4	14.7	392	1	AI284430
43	40	14.6	139	8	N46359
44	39.8	14.5	473	1	AA936680
45	39.8	14.5	528	1	AI435057

#### ALIGNMENTS

RESULT 1

BZ352867

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK\_082724.45.05.x, genomic survey sequence.

ACCESSION

BZ352867

VERSION

BZ352867.1

KEYWORDS

GSS.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

REFERENCE

1 (bases 1 to 228)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,

Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,

Shinn, P., Zimmerman, J., and Ecker, J.R.

A sequence-indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

CONTACT: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA. This sequence lies within 300 bases of the 5' end of

AT3925882.

Class: TDNA tagged.

Location/Qualifiers

1. .228

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Col-0"

/db\_xref="taxon:3702"

/clone="SALK\_082724.45.05.x"

/note="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

BZ352867 228 bp DNA linear GSS 14-NOV-2002

SALK\_082724.45.05.x Arabidopsis thaliana TDNA insertion lines

Arabidopsis thaliana genomic clone SALK\_082724.45.05.x, genomic

survey sequence.

BZ352867 GI:24943729

BZ352867.1

GSS.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 228)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,

Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,

Shinn, P., Zimmerman, J., and Ecker, J.R.

A sequence-indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA. This sequence lies within 300 bases of the 5' end of

AT3925882.

Class: TDNA tagged.

Location/Qualifiers

1. .228

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Col-0"

/db\_xref="taxon:3702"

/clone="SALK\_082724.45.05.x"

/note="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

```

ORIGIN
Query Match      81.0%; Score 222; DB 9; Length 228;
Best Local Similarity 97.4%; Pred. No. 4.3e-46;
Matches 222; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 24 TTTTTCACATGAGCTTCCTTTTATTCMAAGTCAATAAAGTGTGAGGTGATGATCT 83
Db 1 TTTTTCNCNCTGGAGTTCCTTTATTCNNAAGTCNATANAGTGTGAGGTGATGATCT 60

QY 84 TACGGTTTAAACATCGCATGATGTCATTAGCATCAATCCACCGTCCAAATTTAT 143
Db 61 TACGGTTTAAACATCGCATGATGTCATTAGCATCAATCCACCGTCCAAATTTAT 120

QY 144 TAGTTGTTGACAATATCGACCGTCTAAAGTTCACACCGCGCTATAAGAGTTTCATTAT 203
Db 121 TAGTTGTTGACAATATCGACCGTCTAAAGTTCACACCGCGCTATAAGAGTTTCATTAT 180

QY 204 AAATTTTAGCAAAATAAATCGAATAATTTTCTTGACTAAGCT 251
Db 181 AAATTTTAGCAAAATAAATCGAATAATTTTCTTGACTAAGCT 228

RESULT 2
BZ770600      152 bp DNA linear GSS 13-MAR-2003
LOCUS
DEFINITION
SALK 143535.56.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_143535.56.00.x, genomic
survey sequence.
BZ770600
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 152)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At3G25882.
Class: TDNA tagged.
Location/Qualifiers
1..152
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_143535.56.00.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      55.5%; Score 152; DB 9; Length 152;
Best Local Similarity 97.4%; Pred. No. 4.3e-46;
Matches 152; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Best Local Similarity 100.0%; Pred. No. 2.9e-28;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 GCATGATGATGATAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAAAT 159
Db 1 GCATGATGATGATAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGACAAAT 60

QY 160 CGACCGTCTAAGTTCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAAAATA 219
Db 61 CGACCGTCTAAGTTCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAAAATA 120

QY 220 AAATCAGCAAAATAATTTTCTTGACTAAGCT 251
Db 121 AAATCAGCAAAATAATTTTCTTGACTAAGCT 152

RESULT 3
BZ814587      237 bp DNA linear GSS 02-MAY-2002
LOCUS
DEFINITION
SALK 066674 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_066674, genomic survey sequence.
BZ814587
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 237)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..237
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_066674"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

FEATURES
source
1..237
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_066674"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      38.8%; Score 106.2; DB 9; Length 237;
Best Local Similarity 93.3%; Pred. No. 1.4e-16;
Matches 111; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 156 ATATGACCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAA 215
Db 1 ATATGACCGTCTAAGTTCCACACCGCGCTATAAGAGTTTCATTATAAATTTTAGCAA 60

QY 216 AATAAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGCGCGTTAATTTTC 274
Db 61 AATAAAATCAGCAAAATAATTTTCTTGACTAAGCTGAAGCGGCGGTAAACATTTCTC 119

```

RESULT 4	BZ352865	147 bp	DNA	linear	GSS 14-NOV-2002
LOCUS	SALK_082722.32.45.x Arabidopsis thaliana	TDNA insertion lines			
DEFINITION	Arabidopsis thaliana genomic clone SALK_082722.32.45.x, genomic survey sequence.				
ACCESSION	BZ352865				
VERSION	BZ352865.1	GI:24943727			
KEYWORDS	GSS.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi				
REFERENCE	1 (bases 1 to 147)				
AUTHORS	Alonso,J.M., Leishe,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.				
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of AT3G25882. Class: TDNA tagged.				
FEATURES	Location/Qualifiers				
source	1..147				
	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				
	/ecotype="Col-0"				
	/db_xref="taxon:3702"				
	/clone="SALK_082722.32.45.x"				
	/clone_lib="Arabidopsis thaliana TDNA insertion lines"				
	/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <a href="http://signal.salk.edu/tdna_protocols.html">http://signal.salk.edu/tdna_protocols.html</a> "				
ORIGIN					
Query Match	36.9%;	Score 101.2;	DB 9;	Length 147;	
Best Local Similarity	92.7%;	Pred. No. 2.6e-15;			
Matches 139;	Conservative 0;	Mismatches 8;	Indels 3;	Gaps 3;	
Qy	102	ATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTAGTTGTTGACAAATATCG	161		
Db	1	ATGATGATGGTATTAGCTTGAATCTCCCTCTCCAAATTT-TTTACTTGTGA-AATATCG	58		
Qy	162	ACCGTCTAAGTTCCACACCGCGGTATAGAGTTTCATTATAATTTTACCAAAATAAA	221		
Db	59	ACTGCTAAGTTCCAC-CCGACGGCTATAGAGTTTCATTATAATTTTACCAAAATAAA	117		
Qy	222	ATCAGCAATAATTTTTTCTTGACTTAAGCT	251		
Db	118	ATCAGCAATAATTTTTTCTTGACTTAAGCT	147		
RESULT 5	BZ019485	691 bp	DNA	linear	GSS 08-OCT-2002
LOCUS	BZ019485				
DEFINITION	oeg72905.g1 B.oleracea002 Brassica oleracea genomic, genomic survey sequence.				
ACCESSION	BZ019485				

Iida, J., Kawai, J., Sasaki, D., Shiraki, T., Hayashizaki, Y. and Shinozaki, K.  
Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs  
Unpublished (2005)  
Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@tc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998, 2002). This clone is in a modified pBluescript vector.  
Please visit our web site (<http://pfweb.gsc.riken.jp>) and <http://raige.gsc.riken.jp> for further details.

Location/Qualifiers  
1..397  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAFL21-66-M11"  
/lab\_host="DH10B"  
/clone\_lib="RAFL21"  
/note="Site 1: BamHI; Site 2: SalI; Subtraction Library.  
The sequence was obtained from samples subjected to various stress and plant hormones-treated"

ORIGIN

Query Match 17.9%; Score 49; DB 3; Length 397;  
Best Local Similarity 100.0%; Pred. No. 0.056;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAAACATTTTC 274  
DB 1 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAAACATTTTC 49

RESULT 7  
B52379/c  
LOCUS B52379  
DEFINITION CIT-HSP-2003E3.TP CIT-HSP Homo sapiens genomic clone 2003E3,  
genomic survey sequence.  
ACCESSION B52379  
VERSION B52379.1 GI:2606713  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 465)  
Adams, M.D., Rounale, S.D., Field, C.E., Bass, S., Linher, K.,  
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
Simon, M. and Venter, J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building  
Unpublished (1997)  
Other\_GSSs: CIT-HSP-2003E3.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M33-21  
Class: BAC ends.  
Location/Qualifiers  
1..465

FEATURES  
source

Ida, J., Kawai, J., Sasaki, D., Shiraki, T., Hayashizaki, Y. and Shinozaki, K.  
Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs  
Unpublished (2005)  
Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@tc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998, 2002). This clone is in a modified pBluescript vector.  
Please visit our web site (<http://pfweb.gsc.riken.jp>) and <http://raige.gsc.riken.jp> for further details.

Location/Qualifiers  
1..397  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAFL21-66-M11"  
/lab\_host="DH10B"  
/clone\_lib="RAFL21"  
/note="Site 1: BamHI; Site 2: SalI; Subtraction Library.  
The sequence was obtained from samples subjected to various stress and plant hormones-treated"

ORIGIN

Query Match 17.9%; Score 49; DB 3; Length 397;  
Best Local Similarity 100.0%; Pred. No. 0.056;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAAACATTTTC 274  
DB 1 GCAATAATTTTCTTGACTAGCTTAACGACGCCGTTAAACATTTTC 49

RESULT 7  
B52379/c  
LOCUS B52379  
DEFINITION CIT-HSP-2003E3.TP CIT-HSP Homo sapiens genomic clone 2003E3,  
genomic survey sequence.  
ACCESSION B52379  
VERSION B52379.1 GI:2606713  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 465)  
Adams, M.D., Rounale, S.D., Field, C.E., Bass, S., Linher, K.,  
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
Simon, M. and Venter, J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building  
Unpublished (1997)  
Other\_GSSs: CIT-HSP-2003E3.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M33-21  
Class: BAC ends.  
Location/Qualifiers  
1..465

FEATURES  
source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7038437"  
/db\_xref="taxon:9606"  
/clone="2003E3"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII"

ORIGIN

Query Match 16.6%; Score 45.4; DB 9; Length 465;  
Best Local Similarity 56.3%; Pred. No. 0.47;  
Matches 85; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 103 TGATGATGTCATTAGCATCAATCTCCACGCTCCAAATTTATTATTAGTTGTGCAATATCA 162  
DB 388 TAATGAAGTAGGAGCAAGATTATCCATTGAAGGTATTATTATATGCAGCTCACTTAA 329  
QY 163 CGCTTAAGTTCCACACCGAGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAA 222  
DB 328 GTGACAAAATTCATACAGAGACTATAACAGAAATCATATTATAATATTAAATTAAT 269  
QY 223 TCAGCAAAATTAATTTTCTTGACTAAGCTTA 253  
DB 268 ACTTCAATATCTTTTTCACATTAAGATGATTA 238

RESULT 8  
AI990512  
LOCUS AI990512  
DEFINITION ws40e09.x1 NCI CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2499688 3',  
mRNA sequence.  
ACCESSION AI990512  
VERSION AI990512.1 GI:5837393  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 609)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 444.  
Location/Qualifiers  
1..609  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2499688"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP GC6"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA  
from the normalized library NCI-CGAP GC4 was prepared, and  
ss circles were made in vitro. Following HAP purification,

FEATURES  
source



this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN
Query Match      15.9%; Score 43.6; DB 1; Length 609;
Best Local Similarity 58.5%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 124 TCTCCACCGTCCCAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACCGAC 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 TTTCCATTGTAGTGATTATTTATTTTGCAGCTCATCTTAAGTGACAAAATCCATACAGAA 62

QY 184 GGCATATAAGGTTTCATTTAATATTTTTCAGCAAAATAAATCAGCAAAATATTTTCTTG 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 GACTATAACAGAAATCATATTTAATATTTAATATTTAATAATTAATCTTCAATATCTTTCACAT 122

QY 244 ACTAAGCTTA 253
      ||||| |||||
Db 123 AGATGATTA 132

```

```

RESULT 9
R88734
LOCUS
DEFINITION
YP93b07.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:194965 3', mRNA sequence.
ACCESSION
R88734
VERSION
R88734.1 GI:953561
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 321)
Hillier,B., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Contact: Wilton RK
Unpublished (1995)
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 692
High quality sequence stops: 260
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 692 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 260.
Location/Qualifiers
1..321
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3764015"
/db_xref="taxon:9606"
/clone="IWAGE:194965"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer

```

```

FEATURES
source
1..321
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3764015"
/db_xref="taxon:9606"
/clone="IWAGE:194965"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer

```

[5' AACTGGAGAAATTAATAAGATCTTTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN
Query Match      15.7%; Score 43; DB 8; Length 321;
Best Local Similarity 56.7%; Pred. No. 1.9; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 120 TCAATCTCCACCGTCCAAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACAC 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 TNANATTCATTTGAAGGTATTTATTTATTTGCGAGCTCATCTTAAGTGACAAAATTCATAC 63

QY 180 CGACGCTATAGAGTTTCATTTAATAATTTTACGCAAAATAAATCAGCAAAATATTTT 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 AGAAGACTATACAGAAATCATATTTATATATTTAATAATTAATCTTCAATATCTTTCA 123

QY 240 CTTGACTAAGCTTA 253
      ||||| |||||
Db 124 CATTANGATTA 137

```

```

RESULT 10
CNS016E2
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106628
VERSION
AL106628.1 GI:5622852
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1204)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.
Location/Qualifiers
1..1204
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15A12"
/clone_lib="DrosBAC"
/plasmid="pbeloBAC11"
/notes="end : T7"

```

```

COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.

```

```

FEATURES
source
1..1204
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15A12"
/clone_lib="DrosBAC"
/plasmid="pbeloBAC11"
/notes="end : T7"

```

```

ORIGIN
Query Match      15.4%; Score 42.2; DB 10; Length 1204;
Best Local Similarity 16.9%; Pred. No. 3.3; Indels 0; Gaps 0;
Matches 43; Conservative 117; Mismatches 95; Indels 0; Gaps 0;

QY 19 ACCAATTTTCCACATGACGCTTCTTTATTTCCAAAGTCAATAAGTGTGACGTCATG 78
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 830 MMCMTTCTTCTCTTMMMMMMMMTTHHMMMMTTHHMMMMTTHHMMMMTTHHMMMMTTHHMM 889
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 ATACTTACGCTTTTAAACATCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2006, 22:55:15 ; Search time 498 Seconds  
(without alignments)  
3666.919 Million cell updates/sec

Title: US-10-800-161-28  
Perfect score: 274  
Sequence: 1 tctagaatatagcgatcac.....acgacgcggttaacattttc 274

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	274	100.0	274	13	Adri6847 Arabidops
2	272.4	99.4	274	13	Adri6845 Arabidops
3	272.4	99.4	544	13	Adri6846 Arabidops
4	272.4	99.4	862	13	Adri6844 Arabidops
5	272.4	99.4	862	14	Adv66225 Arabidops
6	272.4	99.4	962	13	Adri6843 Arabidops
7	272.4	99.4	1700	12	Adri6822 Arabidops
8	272.4	99.4	1700	13	Adri6822 Arabidops
9	272.4	99.4	1700	14	Adv66204 Arabidops
10	258.4	94.3	1226	14	Adw12661 A. thalia
11	52.8	19.3	509	12	Adri6820 Arabidops
12	52.8	19.3	509	13	Adri6820 Arabidops
13	52.8	19.3	509	14	Adv66202 Arabidops
14	43	15.7	321	6	Abz35565 Human gen
15	42	15.3	229	4	Aas25538 Human ova
16	42	15.3	1197	3	Aas2606 Eosinophi
17	42	15.3	2000	13	Adp5764 Human PRO
18	42	15.3	4641	8	Abz35897 Human sec
19	42	15.3	4673	6	Abz35431 Human gen

c	20	42	15.3	4779	10	ADF90715
c	21	41.4	15.1	2026	6	ABV77883
	22	39.4	14.4	375	9	ACH50239
	23	38.8	14.2	6301	6	ABL32053
	24	38.2	13.9	2000	11	ACL35363
	25	37.8	13.8	319608	3	AAS15601
	26	37.8	13.8	319608	5	AAS09301
	27	37.2	13.6	1834	14	ADW26070
	28	37.2	13.6	1834	14	ADW26105
	29	37.2	13.6	1852	10	ADD22878
	30	37.2	13.6	1963	10	ADD22877
	31	37.2	13.6	1969	14	ADW26073
	32	37.2	13.6	1969	14	ADW26106
c	33	37.2	13.6	2000	11	ACL37108
	34	36	13.1	182328	12	ADL08128
	35	35.6	13.0	87394	13	ADT55151
	36	35.4	12.9	478	13	ACF91349
c	37	35.4	12.9	589	4	AAR80413
	38	35.4	12.9	140036	6	AAS98600
c	39	35.2	12.8	621	8	ACA23093
	40	35.2	12.8	110000	2	AAX20248_04
	41	35	12.8	516	12	ACH68922
c	42	35	12.8	2000	11	ACL35887
	43	34.8	12.7	2208	12	ADO35476
	44	34.8	12.7	2208	13	ADS95227
	45	34.8	12.7	2208	13	ADV68084

ALIGNMENTS

RESULT 1

ADRI6847

ID ADRI6847 standard; DNA; 274 BP.

XX ADRI6847;

DT 21-OCT-2004 (first entry)

DE Arabidopsis thaliana NI16 promoter DNA #5.

XX NI16; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.

XX Arabidopsis thaliana.

PN US2004154051-A1.

XX 05-AUG-2004.

XX 12-MAR-2004; 2004US-00800161.

XX 15-DEC-1999; 99US-0171008P.

PR 11-JAN-2000; 2000US-0175519P.

PR 08-DEC-2000; 2000US-0073368S.

PR 20-JAN-2004; 2004US-00760752.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Cade RM, Dietrich RA, Lawton KA;

XX WPI; 2004-580223/56.

PT Novel isolated nucleic acid molecule comprising Arabidopsis NI16

PT promoter, useful in regulating transcription of coding sequence of

XX interest.

PS Claim 1; SEQ ID NO 28; 39pp; English.

XX The present invention relates to an Arabidopsis NI16 gene promoter useful  
CC in regulating transcription of coding sequence of interest. The invention  
CC is useful in production of transgenic plant or seed exhibiting resistance  
CC to herbicides and microorganism such as Phytophthora parasitica,  
CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The

CC present sequence is *Arabidopsis thaliana* N116 promoter DNA. This sequence  
CC is used in the exemplification of the invention.

XX	SQ	Sequence	274 BP; 91 A; 58 C; 36 G; 89 T; 0 U; 0 Other;
		Query Match	100.0%; Score 274; DB 13; Length 274;
		Best Local Similarity	100.0%; Pred. No. 5.2e-64;
		Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1	TCTAGAAATATAGCGGATACCAATTTTTCACACATGCACTTCCTTTATTCCTAAAGTCA	60
Db	1	TCTAGAAATATAGCGGATACCAATTTTTCACACATGCACTTCCTTTATTCCTAAAGTCA	60
QY	61	ATAAAGTGTGACGGTCATGATACTTTACGGCTTTTAAACATCGCATGATGATGTCATTAGCAT	120
Db	61	ATAAAGTGTGACGTCATGATACTTTACGGCTTTTAAACATCGCATGATGATGTCATTAGCAT	120
QY	121	CAATCTCCACCGTCCAAATTTATTTAGTTGTTGTGACAATATCGACCGCTCTAAGTTCACACC	180
Db	121	CAATCTCCACCGTCCAAATTTATTTAGTTGTTGTGACAATATCGACCGCTCTAAGTTCACACC	180
QY	181	GACGGCTATAAGAGTTTCATTATATAATTTTATAGCAAAATAAAATCAGCAAAATAATTTTTTTC	240
Db	181	GACGGCTATAAGAGTTTCATTATATAATTTTATAGCAAAATAAAATCAGCAAAATAATTTTTTTC	240
QY	241	TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC	274
Db	241	TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC	274

RESULT 2  
ADRI6845  
ID ADRI6845 standard; DNA; 274 BP.  
XX  
AC ADRI6845;  
XX  
21-OCT-2004 (first entry)  
XX  
XX  
DE DE  
XX  
XX Arabidopsis thaliana N116 promoter DNA #3.  
XX N116; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.

CC	is used in the exemplification of the invention.					
XX						
SQ	Sequence	274 BP; 92 A; 58 C; 35 G; 89 T; 0 U; 0 Other;				
	Query Match	99.4%; Score 272.4; DB 13; Length 274;				
	Best Local Similarity	99.6%; Pred. No. 1.4e-63;				
	Matches 273; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
Qy	1	TCTAGAAATATAGCGGATACCAATTTTTCACACACATGGACTTCCTTTATTCCAAAAGTCA	60			
Dd	1	TCTAAAAATATAGCGGATACCAATTTTTCACACATGGACTTCCTTTATTCCAAAAGTCA	60			
Qy	61	ATAAAGTGTCAGGTGCATGATACCTTTTAAACAATCGCATGATGATGTCATTAGCAT	120			
Dd	61	ATAAAGTGTCAGGTGCATGATACCTTTTAAACAATCGCATGATGATGTCATTAGCAT	120			
Qy	121	CAATCTCCACCGGCCAATTTTATTAGTGTGTGAACAATATCACCAGTCTTAAGTTCACACC	180			
Dd	121	CAATCTCCACCGGCCAATTTTATTAGTGTGTGAACAATATCACCAGTCTTAAGTTCACACC	180			
Qy	181	GACGGCTATAAGAGTTTTTCATTATAAAATTTTAGCAAAAATAAAATCAGCAAATAAATTTTTTTC	240			
Dd	181	GACGGCTATAAGAGTTTTTCATTATAAAATTTTAGCAAAAATAAAATCAGCAAATAAATTTTTTTC	240			
Qy	241	TTGACTAAGCTTAAACGACGCCGTTAACATTTTC	274			
Dd	241	TTGACTAAGCTTAAACGACGCCGTTAACATTTTC	274			

XX	RESULT 3
ADRI6846	
ID	ADRI6846 standard; DNA; 544 BP.
XX	
AC	ADRI6846;
XX	
DT	21-OCT-2004 (first entry)
XX	
DE	Arabidopsis thaliana NIL6 promoter DNA #4.
XX	
KW	NIL6; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	US2004154051-A1.
XX	
PD	05-AUG-2004.
XX	
PF	12-MAR-2004; 2004US-00800161.
XX	
PR	15-DEC-1999; 99US-0171008P.
PR	11-JAN-2000; 2000US-0175519P.
PR	08-DEC-2000; 2000US-00733685.
PR	20-JAN-2004; 2004US-00760752.
XX	
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Cade RM, Dietrich RA, Lawton KA;
XX	
DR	WPI; 2004-580223/56.
XX	
PT	Novel isolated nucleic acid molecule comprising Arabidopsis NIL6
PT	promoter, useful in regulating transcription of coding sequence of
PT	interest.
XX	
PS	Claim 12; SEQ ID NO 27; 39pp; English.
XX	
CC	The present invention relates to an Arabidopsis NIL6 gene promoter useful
CC	in regulating transcription of coding sequence of interest. The invention
CC	is useful in production of transgenic plant or seed exhibiting resistance
CC	to herbicides and microorganism such as Phytophthora parasitica,
CC	Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
CC	present sequence is Arabidopsis thaliana NIL6 promoter DNA. This sequence
CC	is used in the exemplification of the invention.

XX SQ Sequence 544 BP; 209 A; 84 C; 95 G; 156 T; 0 U; 0 Other;  
Query Match 99.4%; Score 272.4; DB 13; Length 544;  
Best Local Similarity 99.6%; Pred. No. 1.6e-63;  
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAAGTCA 60  
DB 271 TCTAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAAGTCA 330  
QY 61 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120  
DB 331 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 390  
QY 121 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACC 180  
DB 391 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACC 450  
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 240  
DB 451 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 510  
QY 241 TTGACTAAGCTTAAACGACCGCGTTAAACATTTTC 274  
DB 511 TTGACTAAGCTTAAACGACCGCGTTAAACATTTTC 544

RESULT 4  
ADRI6844  
ID ADRI6844 standard; DNA; 862 BP.  
XX AC ADRI6844;  
XX DT 21-OCT-2004 (first entry)  
XX DE Arabidopsis thaliana N116 promoter DNA #2.  
XX KW N116; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.  
XX OS Arabidopsis thaliana.

XX PN US2004154051-A1.  
XX PD 05-AUG-2004.  
XX PF 12-MAR-2004; 2004US-00800161.  
XX PR 15-DEC-1999; 99US-0171008P.  
XX PR 11-JAN-2000; 2000US-0175519P.  
XX PR 08-DEC-2000; 2000US-00733685.  
XX PR 20-JAN-2004; 2004US-00760752.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Cade RM, Dietrich RA, Lawton KA;  
XX WPI; 2004-580223/56.  
XX Novel isolated nucleic acid molecule comprising Arabidopsis N116  
XX promoter, useful in regulating transcription of coding sequence of  
XX interest.  
XX Claim 12; SEQ ID NO 25; 39pp; English.

XX The present invention relates to an Arabidopsis N116 gene promoter useful  
XX in regulating transcription of coding sequence of interest. The invention  
XX is useful in production of transgenic plant or seed exhibiting resistance  
XX to herbicides and microorganisms such as Phytophthora parasitica,  
XX pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
XX present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence  
XX is used in the exemplification of the invention.

SQ Sequence 862 BP; 337 A; 112 C; 152 G; 261 T; 0 U; 0 Other;  
Query Match 99.4%; Score 272.4; DB 13; Length 862;  
Best Local Similarity 99.6%; Pred. No. 1.8e-63;  
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAAGTCA 60  
DB 578 TCTAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAAGTCA 637  
QY 61 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 120  
DB 638 ATAAAGTGTGACGTCAATGATCTTACGCTTTTAAACATCGCATGATGTCATTAGCAT 697  
QY 121 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACC 180  
DB 698 CAATCTCCACCGTCCAAATTTATTAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACC 757  
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 240  
DB 758 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 817  
QY 241 TTGACTAAGCTTAAACGACCGCGTTAAACATTTTC 274  
DB 818 TTGACTAAGCTTAAACGACCGCGTTAAACATTTTC 851

RESULT 5  
ADV66225  
ID ADV66225 standard; DNA; 862 BP.  
XX AC ADV66225;  
XX DT 24-FEB-2005 (first entry)  
XX DE Arabidopsis thaliana N116 promoter.  
XX KW Systemic acquired resistance; disease-resistance; transgenic plant;  
XX plant; N116; N1M1 interactor; promoter; ds.  
XX OS Arabidopsis thaliana.  
XX Key Location/Qualifiers  
XX FT promoter 1..862  
XX FT /tag= a  
XX FT /note= "This promoter is specifically claimed"  
XX FT misc\_feature 365..374  
XX FT /tag= b  
XX FT /note= "TCAI motif"  
XX FT misc\_feature 426..435  
XX FT /tag= c  
XX FT /note= "TCAI motif"  
XX FT enhancer 609..614  
XX FT /tag= d  
XX FT /note= "MYCATR22 element"  
XX FT enhancer 646..665  
XX FT /tag= e  
XX FT /note= "CMV AS1 salicylic acid response element"  
XX FT enhancer 707..712  
XX FT /tag= f  
XX FT /note= "PAL BOX"  
XX FT enhancer 757..762  
XX FT /tag= g  
XX FT /note= "HEXAMERAT 4 element"  
XX PN US2004248303-A1.  
XX 09-DEC-2004.  
XX 20-JAN-2004; 2004US-00760752.  
XX 15-DEC-1999; 99US-0171008P.  
XX 11-JAN-2000; 2000US-0175519P.

```

PR 08-DEC-2000; 2000US-00733685.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Cade RM, Dietrich RA;
XX WPI; 2005-020569/02.
DR P-PSDB; ADV66203.
XX
XX Novel isolated N116 nucleic acid promoter, useful for initiating
PT transcription of DNA that encodes proteins involved in regulation of
PT systemic acquired resistance (SAR) gene expression in plants.
XX
XX Claim 1; Page; 32pp; English.
XX
XX The invention relates to a N116 nucleic acid promoter sequence. The
CC promoter sequence of the invention is useful for initiating transcription
CC of DNA that encodes proteins involved in regulation of systemic acquired
CC resistance (SAR) gene expression in plants. It enhances the expression of
CC SAR genes such as pathogenesis-related protein (PR-1), thus allows broad-
CC spectrum disease resistance in plant such as resistance against
CC Pseudomonas syringae. The present sequence is the Arabidopsis thaliana
CC N116 gene including 5' upstream promoter. Note: This sequence is not
CC given separately, but has been created from the N116 gene (ADV66204).
XX given in the sequence listing and the information provided in claim 1.
XX
SQ Sequence 862 BP; 337 A; 112 C; 152 G; 261 T; 0 U; 0 Other;

Query Match 99.4%; Score 272.4; DB 14; Length 862;
Best Local Similarity 99.6%; Pred. No. 1.8e-63;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
Db 578 TCTAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 637
QY 61 ATAAAGTGTGAGTCATGATCTTACGCTTTAAACATCGCATGATGATGATGATGAT 120
Db 638 ATAAAGTGTGAGTCATGATCTTACGCTTTAAACATCGCATGATGATGATGATGAT 697
QY 121 CAATCTCCACCGTCCCAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACACC 180
Db 698 CAATCTCCACCGTCCCAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACACC 757
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
Db 758 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 817
QY 241 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 274
Db 818 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 851

RESULT 6
ADRI6843
ID ADRI6843 standard; DNA; 962 BP.
XX
XX AC ADRI6843;
XX
XX DT 21-OCT-2004 (first entry)
XX
XX DE Arabidopsis thaliana N116 promoter DNA #1.
XX
XX KW N116; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.
XX
XX OS Arabidopsis thaliana.
XX
XX PN US2004154051-A1.
XX
XX PD 05-AUG-2004.
XX
XX PF 12-MAR-2004; 2004US-00800161.
XX

```

```

PR 15-DEC-1999; 99US-0171008P.
PR 11-JAN-2000; 2000US-0175519P.
PR 08-DEC-2000; 2000US-00733685.
PR 20-JAN-2004; 2004US-00760752.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cade RM, Dietrich RA, Lawton KA;
XX
XX WPI; 2004-580223/56.
XX
XX Novel isolated nucleic acid molecule comprising Arabidopsis N116
PT promoter, useful in regulating transcription of coding sequence of
PT interest.
XX
XX Claim 11; SEQ ID NO 24; 39pp; English.
XX
XX The present invention relates to an Arabidopsis N116 gene promoter useful
CC in regulating transcription of coding sequence of interest. The invention
CC is useful in production of transgenic plant or seed exhibiting resistance
CC to herbicides and microorganism such as Phytophthora parasitica,
CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence
CC is used in the exemplification of the invention.
XX
XX Sequence 962 BP; 379 A; 125 C; 165 G; 293 T; 0 U; 0 Other;

Query Match 99.4%; Score 272.4; DB 13; Length 962;
Best Local Similarity 99.6%; Pred. No. 1.8e-63;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 60
Db 678 TCTAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCAAAGTCA 737
QY 61 ATAAAGTGTGAGTCATGATCTTACGCTTTAAACATCGCATGATGATGATGATGAT 120
Db 738 ATAAAGTGTGAGTCATGATCTTACGCTTTAAACATCGCATGATGATGATGATGAT 797
QY 121 CAATCTCCACCGTCCCAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACACC 180
Db 798 CAATCTCCACCGTCCCAATTTATTTAGTTGTGACAAATATCGACCGTCTAAGTTCCACACC 857
QY 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240
Db 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917
QY 241 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 274
Db 918 TTGACTAAGCTTAAACGACGCGGTTAAACATTTTC 951

RESULT 7
ADNI7259
ID ADNI7259 standard; DNA; 1700 BP.
XX
XX AC ADNI7259;
XX
XX DT 17-JUN-2004 (first entry)
XX
XX DE Arabidopsis thaliana N116 gene.
XX
XX KW N116; SAR; systemic acquired resistance; mouse-ear cress; gene; ds.
XX
XX OS Arabidopsis thaliana.
XX
XX FH Key Location/Qualifiers
FT misc_feature 365..374
FT /*tag= a
FT /*note= "TCAL motif"
FT misc_feature 426..435
FT /*tag= b
FT /*note= "TCAL motif"
FT

```



```

FT misc_feature 609..614
FT /tag= c
FT /note= "MYCATR22 element"
FT misc_feature 646..665
FT /tag= d
FT /note= "CAMV AS1 salicylic acid response element"
FT misc_feature 707..712
FT /tag= e
FT /note= "PAL box"
FT misc_feature 757..762
FT /tag= f
FT /note= "HEXAMERAT 4 element"
FT CDS 863..1231
FT /tag= g
FT /product= "Arabidopsis thaliana N116 protein"
XX
XX US6706952-B1.
XX
XX 16-MAR-2004.
XX
XX 08-DEC-2000; 2000US-00733685.
XX
XX 15-DEC-1999; 99US-0171008P.
XX 11-JAN-2000; 2000US-0175519P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cad RM, Dietrich RA;
XX
XX WPI; 2004-313378/29.
XX P-PSDB; ADN17258.
XX
XX New N116 nucleic acid sequence and encoded protein, useful for increasing
XX systemic acquired resistance gene expression in a plant.
XX
XX Example 2; SEQ ID NO 3; 29pp; English.
XX
XX The invention relates to Arabidopsis N116 gene encoding a protein
XX involved in the regulation of SAR gene expression in plants. The N116
XX nucleic acid molecule and the encoded protein is useful in increasing
XX systemic acquired resistance (SAR) gene expression in a plant. The
XX present sequence is Arabidopsis thaliana N116 gene.
XX
XX Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
XX
XX Query Match 99.4%; Score 272.4; DB 12; Length 1700;
XX Best Local Similarity 99.6%; Pred. No. 2e-63;
XX Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCTCCAAAGTCA 60
Db |||||
578 TCTAAAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCTCCAAAGTCA 637
Qy |||||
61 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
Db |||||
638 ATAAAGTGTGACGTCATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 697
Qy |||||
121 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGCACAATATCGACGCTCTAAGTTCACACC 180
Db |||||
698 CAATCTCCACCGTCCAAATTTATTTAGTTGTTGCACAATATCGACGCTCTAAGTTCACACC 757
Qy |||||
181 GACGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 240
Db |||||
758 GACGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAATTTTTC 817
Qy |||||
241 TTGACTAGCTTAACGACGCGGTTAACAATTTTC 274
Db |||||
818 TTGACTAGCTTAACGACGCGGTTAACAATTTTC 851

RESULT 8
ADN16822
ID ADN16822 standard; DNA; 1700 BP.

```

---

```

XX AC ADR16822;
XX 21-OCT-2004 (first entry)
XX Arabidopsis thaliana N116 genomic DNA.
XX
XX N116; transgenic; herbicide resistance; mouse-ear cross; ds; gene.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX misc_feature 365..374
XX /tag= a
XX /note= "TCA1 motif"
XX misc_feature 426..435
XX /tag= b
XX /note= "TCA1 motif"
XX misc_feature 609..614
XX /tag= c
XX /note= "MYCATR22 element"
XX misc_feature 646..665
XX /tag= d
XX /note= "CAMV AS1 salicylic acid response element"
XX misc_feature 707..712
XX /tag= e
XX /note= "PAL BOX"
XX misc_feature 757..762
XX /tag= f
XX /note= "HEXAMERAT 4 element"
XX CDS 863..1231
XX /tag= g
XX /product= "Arabidopsis thaliana N116 protein"
XX
XX US2004154051-A1.
XX
XX 05-AUG-2004.
XX
XX 12-MAR-2004; 2004US-00800161.
XX
XX 15-DEC-1999; 99US-0171008P.
XX 11-JAN-2000; 2000US-0175519P.
XX 08-DEC-2000; 2000US-00733685.
XX 20-JAN-2004; 2004US-00760752.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cade RM, Dietrich RA, Lawton KA;
XX
XX WPI; 2004-580223/56.
XX P-PSDB; ADR16821.
XX
XX Novel isolated nucleic acid molecule comprising Arabidopsis N116
XX promoter, useful in regulating transcription of coding sequence of
XX interest.
XX
XX Example 3; SEQ ID NO 3; 39pp; English.
XX
XX The present invention relates to an Arabidopsis N116 gene promoter useful
XX in regulating transcription of coding sequence of interest. The invention
XX is useful in production of transgenic plant or seed exhibiting resistance
XX to herbicides and microorganism such as Phytophthora parasitica,
XX Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The
XX present sequence is Arabidopsis thaliana N116 genomic DNA, including the
XX 5'upstream promoter sequence.
XX
XX Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
XX
XX Query Match 99.4%; Score 272.4; DB 13; Length 1700;
XX Best Local Similarity 99.6%; Pred. No. 2e-63;
XX Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
Qy 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATTCCTCCAAAGTCA 60

```

Db 578 TCTAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATCCAAAGTCA 637  
 QY 61 ATAAAGTGTGAGTCATGATATCTAGCGCTTTAAACATCGCATGATGATGTCATTAGCAT 120  
 Db 638 ATAAAGTGTGAGTCATGATATCTAGCGCTTTAAACATCGCATGATGATGTCATTAGCAT 697  
 QY 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTCACATATCGACCGTCTTAAGTTCCACACC 180  
 Db 698 CAATCTCCACCGTCCAAATTTATTTAGTTGTCACATATCGACCGTCTTAAGTTCCACACC 757  
 QY 181 GACGGCTATAAGAGTTTCAATTAATAATTTAGCAAAATAAAATCAGCAATAATATTTTTC 240  
 Db 758 GACGGCTATAAGAGTTTCAATTAATAATTTAGCAAAATAAAATCAGCAATAATATTTTTC 817  
 QY 241 TTGACTAAGCTTAAACGACCGCTTAAACATTTTC 274  
 Db 818 TTGACTAAGCTTAAACGACCGCTTAAACATTTTC 851

RESULT 9  
 ADV66204  
 ID ADV66204 standard; DNA; 1700 BP.

AC ADV66204;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE Arabidopsis thaliana Ni16 gene.  
 XX  
 KW Systemic acquired resistance; disease-resistance; transgenic plant;  
 KW plant; Ni16; NIM1 interactor; gene; ds.  
 XX  
 OS Arabidopsis thaliana.

FH Key Location/Qualifiers  
 FT Promoter 1..862  
 FT /tag= a  
 FT /note= "This promoter is specifically claimed"  
 FT misc\_feature 365..374  
 FT /tag= b  
 FT /note= "TCAT motif"  
 FT misc\_feature 426..435  
 FT /tag= c  
 FT /note= "TCAT motif"  
 FT enhancer 609..614  
 FT /tag= d  
 FT /note= "MYCATR22 element"  
 FT enhancer 646..665  
 FT /tag= e  
 FT /note= "CAMV AS1 salicylic acid response element"  
 FT enhancer 707..712  
 FT /tag= f  
 FT /note= "PAL BOX"  
 FT enhancer 757..762  
 FT /tag= g  
 FT /note= "HEXAMERAT 4 element"  
 FT CDS 863..1231  
 FT /tag= h  
 FT /product= "Arabidopsis thaliana Ni16 protein"

XX US2004248303-A1.  
 PN  
 XX 09-DEC-2004.  
 XX  
 XX 20-JAN-2004; 2004US-00760752.  
 XX  
 PR 15-DEC-1999; 99US-0171008P.  
 PR 11-JAN-2000; 2000US-0175519P.  
 PR 08-DEC-2000; 2000US-0073368S.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

PI Cade RM, Dietrich RA;  
 XX WPI; 2005-020569/02.  
 DR P-PSDB; ADV66203.  
 XX  
 PT Novel isolated Ni16 nucleic acid promoter, useful for initiating  
 PT transcription of DNA that encodes proteins involved in regulation of  
 PT systemic acquired resistance (SAR) gene expression in plants.  
 XX  
 PS Example 3; SEQ ID NO 3; 32pp; English.

XX The invention relates to a Ni16 nucleic acid promoter sequence. The  
 CC promoter sequence of the invention is useful for initiating transcription  
 CC of DNA that encodes proteins involved in regulation of systemic acquired  
 CC resistance (SAR) gene expression in plants. It enhances the expression of  
 CC SAR genes such as pathogenesis-related protein (PR-1), thus allows broad-  
 CC spectrum disease resistance in plant such as resistance against  
 CC Pseudomonas syringae. The present sequence is the Arabidopsis thaliana  
 CC Ni16 gene including 5' upstream promoter. Note: The promoter is  
 CC specifically claimed in claim 1 - the isolated promoter is given in  
 CC ADV66225.  
 XX  
 SQ Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;

Query Match 99.4%; Score 272.4; DB 14; Length 1700;  
 Best Local Similarity 99.6%; Pred. No. 2e-63;  
 Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTAGAAATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATCCAAAGTCA 60  
 Db 578 TCTAAATATATAGCGGATACCAATTTTCCACATGGACTTCCTTTATCCAAAGTCA 637  
 QY 61 ATAAAGTGTGAGTCATGATATCTAGCGCTTTAAACATCGCATGATGATGTCATTAGCAT 120  
 Db 638 ATAAAGTGTGAGTCATGATATCTAGCGCTTTAAACATCGCATGATGATGTCATTAGCAT 697  
 QY 121 CAATCTCCACCGTCCAAATTTATTTAGTTGTCACATATCGACCGTCTTAAGTTCCACACC 180  
 Db 698 CAATCTCCACCGTCCAAATTTATTTAGTTGTCACATATCGACCGTCTTAAGTTCCACACC 757  
 QY 181 GACGGCTATAAGAGTTTCAATTAATAATTTAGCAAAATAAAATCAGCAATAATATTTTTC 240  
 Db 758 GACGGCTATAAGAGTTTCAATTAATAATTTAGCAAAATAAAATCAGCAATAATATTTTTC 817  
 QY 241 TTGACTAAGCTTAAACGACCGCTTAAACATTTTC 274  
 Db 818 TTGACTAAGCTTAAACGACCGCTTAAACATTTTC 851

RESULT 10  
 ADM12661  
 ID ADM12661 standard; DNA; 1226 BP.  
 XX  
 AC ADM12661;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE A. thaliana NIMIN-2 promoter SEQ ID NO:2.  
 XX  
 KW transgenic plant; NIMIN-2; expression; plant; ds.

XX Arabidopsis thaliana.  
 OS  
 XX EP1502953-A1.  
 PN  
 XX 02-FEB-2005.  
 PD  
 XX 28-JUL-2003; 2003EP-00017114.  
 PF  
 XX 28-JUL-2003; 2003EP-00017114.  
 PR  
 XX (UYHO-) UNIV HOHENHEIM.

```
PI Pfizner A, Roth B;
XX WPI; 2005-114423/13.
XX
XX Novel nucleic acid containing selectively inducible NIMIN-1 promoter
PT sequence or NIMIN-2 promoter sequence, useful in controlled expression of
PT desired expression products in suitable host expression organisms such as
PT transgenic plants.
XX
XX Claim 1; SEQ ID NO 2; 22pp; English.
XX
XX The invention relates to a novel nucleic acid (I) containing at least a
CC first nucleotide sequence chosen from a NIMIN-1 promoter sequence of
CC (ADW12660) and NIMIN-2 promoter sequence of (ADW12661), and its
CC biologically active derivative. The polynucleotide of the invention is
CC useful in controlled expression of desired expression products in
CC suitable host expression organisms such as transgenic plants. The present
CC sequence represents the NIMIN-2 promoter.
XX
XX Sequence 1226 BP; 496 A; 146 C; 201 G; 383 T; 0 U; 0 Other;
SQ
Query Match 94.3%; Score 258.4; DB 14; Length 1226;
Best Local Similarity 99.6%; Pred. No. 1.1e-59;
Matches 259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAGTCA 60
Db 967 TCTAAATATATAGCCGATACCAATTTTCCACACATGCGACTTCCTTTATTTCCAAAGTCA 1026
QY 61 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120
Db 1027 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 1086
QY 121 CAATCTCCACCGTCCCAATTTATTAGTTGTGTGACAATATCGACCGTCTAAGTTCACACC 180
Db 1087 CAATCTCCACCGTCCCAATTTATTAGTTGTGTGACAATATCGACCGTCTAAGTTCACACC 1146
QY 181 GACGGCTATAAGAGTTTCATATAATTTTAGCAAAATAAATCAGCAAAATATTTTTC 240
Db 1147 GACGGCTATAAGAGTTTCATATAATTTTAGCAAAATAAATCAGCAAAATATTTTTC 1206
QY 241 TTGACTAAGCTTAAACGACG 260
Db 1207 TTGACTAAGCTTAAACGACG 1226
RESULT 11
ADN17257
ID ADN17257 standard; cDNA; 509 BP.
XX
XX ADN17257;
XX
XX 17-JUN-2004 (first entry)
XX
XX Arabidopsis thaliana NI16 cDNA.
XX
XX NI16; SAR; systemic acquired resistance; mouse-ear cross; gene; ss.
XX
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX CDS 68..436
XX /*tag= a
XX /product= "Arabidopsis thaliana NI16 protein"
XX
XX US6706952-B1.
XX
XX 16-MAR-2004.
XX
XX 08-DEC-2000; 2000US-00733685.
XX
XX 15-DEC-1999; 99US-0171008P.
XX
XX 11-JAN-2000; 2000US-0175519P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cade RM, Dietrich RA, Lawton KA;
XX
XX WPI; 2004-580223/56.
XX
XX P-PSDB; ADN16821.
XX
XX Novel isolated nucleic acid molecule comprising Arabidopsis NI16
PT promoter, useful in regulating transcription of coding sequence of
PT interest.
XX
XX Example 1; SEQ ID NO 1; 39pp; English.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cad RM, Dietrich RA;
XX
XX WPI; 2004-313378/29.
XX
XX P-PSDB; ADN17258.
XX
XX New NI16 nucleic acid sequence and encoded protein, useful for increasing
PT systemic acquired resistance gene expression in a plant.
XX
XX Claim 2; SEQ ID NO 1; 29pp; English.
XX
XX The invention relates to Arabidopsis NI16 gene encoding a protein
CC involved in the regulation of SAR gene expression in plants. The NI16
CC nucleic acid molecule and the encoded protein is useful in increasing
CC systemic acquired resistance (SAR) gene expression in a plant. The
CC present sequence is Arabidopsis thaliana NI16 cDNA.
XX
XX Sequence 509 BP; 158 A; 82 C; 133 G; 136 T; 0 U; 0 Other;
SQ
Query Match 19.3%; Score 52.8; DB 12; Length 509;
Best Local Similarity 96.4%; Pred. No. 0.00021;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 219 AAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGTAAACATTTTC 274
Db 1 AAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGTAAACATTTTC 56
RESULT 12
ADN16820
ID ADN16820 standard; cDNA; 509 BP.
XX
XX ADN16820;
XX
XX 21-OCT-2004 (first entry)
XX
XX Arabidopsis thaliana NI16 cDNA.
XX
XX NI16; transgenic; herbicide resistance; mouse-ear cross; ss; gene.
XX
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX CDS 68..436
XX /*tag= a
XX /product= "Arabidopsis thaliana NI16 protein"
XX
XX US2004154051-A1.
XX
XX 05-AUG-2004.
XX
XX 12-MAR-2004; 2004US-00800161.
XX
XX 15-DEC-1999; 99US-0171008P.
XX
XX 11-JAN-2000; 2000US-0175519P.
XX
XX 08-DEC-2000; 2000US-00733685.
XX
XX 20-JAN-2004; 2004US-00760752.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Cade RM, Dietrich RA, Lawton KA;
XX
XX WPI; 2004-580223/56.
XX
XX P-PSDB; ADN16821.
XX
XX Novel isolated nucleic acid molecule comprising Arabidopsis NI16
PT promoter, useful in regulating transcription of coding sequence of
PT interest.
XX
XX Example 1; SEQ ID NO 1; 39pp; English.
XX
```



Query Match 15.3%; Score 42; DB 4; Length 229;  
Best Local Similarity 57.7%; Pred. No. 0.15;  
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps

***This Page Blank (uspto)***